On the Persuasive Power of Videogame avatars on Health-related behaviours

Oliver James Clark

PhD 2019

On the persuasive power of videogame avatars on health-related behaviours

Oliver James Clark

A thesis submitted in partial fulfilment of the requirements of

Manchester Metropolitan University for the degree of

Doctor of Philosophy

Faculty of Health, Psychology, and Social Change

Manchester Metropolitan University

2019

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others.

For Audrey, who never doubted for a moment.

Acknowledgements

I am incredibly grateful for having such a fantastic supervisory team; Thank you Jenny Cole, Sarah Grogan, and Niki Ray for the comments, suggestions, and encouragement, and for tolerating my single minded dedication to Markdown and LaTeX; and Kevin Tan for the technical advice over the past three years. It has been a pleasure working with you all.

Thank you to my examiners Professor Phillipa Diedrichs and Professor Marc Jones for thoroughly reading the thesis, providing incredibly helpful feedback, and for an excellent discussion, with Professor Sue Powell, during my Viva.

Thanks to the *Chatistician*, Chelsea Parlett-Pelleriti for the *Reviewer-2 Special* - this was invaluable! I also would like to thank the Technical Officer Gary Dicks for his support with getting my laboratory experiments off the ground and doing a fantastic job with the new labs.

I am very grateful for the funding received from the British Psychological Society and the Society for the Improvement of Psychological Science for conference attendance bursaries, and to the MMU Department of Psychology and the Stress, Health, and Performance research group for providing much needed funding for both conferences and vouchers for participants (particularly Jo Ashby and Marc Jones) .

One of the highlights of the programme was the BBSRC Reproducibility Retreat at Cumberland Lodge, organised and delivered by some awesome people. It was a truly valuable experience. Thanks to everyone who participated, and Chris Chambers, Laura Fortunato, Stan Lazic, Dorothy Bishop for running excellent sessions and providing invaluable advice.

I was fortunate enough to be able to pay a visit to the Red Square Lab at Technical University Eindhoven in September 2018. The experience was amazing, and I am eternally grateful to Daniël Lakens, Anne Scheel, Farid Anvari, and Peder Isager for making me feel welcome and introducing me to some really interesting people.

To my friends and colleagues at the British Psychological Society Postgraduate Affairs Group (PsyPAG). PsyPag has been a hidden gem, and one of my few regrets is not getting involved sooner! I will be continuing my positions until the end of my tenure, but I have met some awesome people through both the conferences and meetings. Particularly the Core Committee past and present, who have always been on the end of a WhatsApp group to share indignation and memes. Thank you Maddi Pownall, Cat Talbot, Olly Robertson, Ben Butterworth, Jammy Stacy, Candie Whittaker, Elaine Coxon, Danielle Paddock, and Chris Robus for being a message away for laughs, advice, indignation, and memes.

Early on in the programme, I attended four workshops on Bayesian Statistics. Mark Andrews and Thom Baguley did a stellar job dishing out Bayesian CoolAid at the Prior Exposure workshops and beyond.

A major source of knowledge, advice, and entertainment has been the various academic podcasts that hosts pour hundreds of hours of unpaid work into. A major shout out to James Heathers and Dan Quintana from the Everything Hertz podcast - it's possible I would have entirely missed the Open Science movement without you. Also, thanks to The Black Goat, Two Psychologists, Four Beers, ReproducibiliTea, and Very Bad Wizards.

There have been a huge amount of inspiring people that have offered coding and stats advice over Twitter, at various journal clubs, and conferences. I am really thankful to Lisa DeBruine, Dan Simpson, Oscar Olvera Astiva, Paul Buerkner, Ed Berry, and everyone who has offered advice guidance and support in person, online, and elsewhere.

A very welcome distraction from videogame avatars has been the various collaborations that I have sparked up over the past few years. Thanks to Andrew Stewart, James Bartlett, Emma Henderson, Yael Benn, and Sam Parsons for keeping my mind active! Also, thanks to Tom Hostler for letting me help with the Open Science Working Group at MMU, and for the various late evening drinking and riff sessions.

The Kinisi-Fit group, and Manchester Taekwondo Warriors have been a major outlet for my physical and mental wellbeing. I am incredibly grateful to Michael Dean O'Reilly, Natalie Constanti, Ross Onufryk, Sarah Barker, and Matt Houston for their advice, sticky stop-watches, and providing the opportunity to hit stuff!

The vast majority of my thesis was written over guest filter coffees and brunch at two local Third Wave Coffee establishments. Thank you to all of the staff at Takk Oxford Road and University Green, and Foundation Whitworth Lock for being patient over the hundreds of hours I must have spent nursing cups of coffee!

My family have been wonderfully supportive, and I'd like to thank Jane Wright, Ian Wright, and Tim Clark for being available for phone calls and videogames. I would like to thank Han Clark for always believing in me.

I am fortunate to have some excellent friends: I've had some great downtime drinking beers and going to gigs with orta Therox, Jon Wingrove, James Twiddy, and Matthew Barrell. Dr Awesome Kev Elner who encouraged me to get back on the Academic Horse, gave great advice on the perils of perfectionism and has been an excellent support over my various academic endeavours.

My lab mates Lucy Walker, Kat Fradley, Rebecca Dwyer, Rachel Zieme, and Georgia Wilson have been a great support during the long hours waiting around for participants.

I'd like thank Dr A Gledhill, who got me over the last hurdle of the thesis.

Last, but not least, thanks to my *El Dude Bro* Sofia Persson for being an amazing friend, colleague, and academic exemplar; it was from early conversations with you that I plucked up the courage to jump back into education.

Abbreviations

Term	Abbreviation
Associated Systems Theory	AST
For what it’s worth	FWIW
If You’re Interested	IYI
Massively Multiplayer Online RPG	MMORPG
Non-playable character	NPC
Open Science Framework	OSF
Preferred Reporting Items for Systematic review and Meta-Analysis	PRISMA
Role Playing Game	RPG
Self-Discrepancy Theory	SDT
Self-Perception Theory	VSM
Virtual Self Modelling	SPT

Glossary

Term	Abbreviation
Affordances	Customisable elements of an avatar, e.g. weight or skin colour
Agent	A computer-controlled character.
Autotelic activity	An activity that is to be enjoyed in and of itself
Avatar	The representation of the self in a virtual environment
Chocobo	A bird-like creature from the Final Fantasy series
Exemplar	A non-controlled representation of a group
Exergame	A videogame that uses exercise-like actions as input.
Non-Playable Character	Another word for agent
Serious games	Games with a message, or educational purpose outside of enjoyment
Shovelware	Hastily hacked software or games bundled to imply value
Tank	A videogame character whose main purpose is to take damage.

Abstract

Background: Avatars are representations of the self in a virtual environment. They have been used to influence behaviour and may represent a promising avenue for designing interventions to promote health-related behaviour change.

Aim: To determine the extent to which a representation of the self in a videogame influences health-related behaviours.

Method: In addressing the aim of the thesis, a mixed methods approach was adopted. This started with a systematic review investigating the effectiveness of various attempts at health persuasion using avatar appearance manipulation. Next, a qualitative study investigating gamers' avatar design preferences and experiences of playing an exergame with an idealised, self-similar avatar was conducted. The quantitative phase of the thesis involved three quantitative studies investigating the existence, ambivalence, and variance of stereotypes associated with plus-sized and athletic physiques over three modalities (text, image, video); and a replication-extension of an exergame-based Proteus Effect study involving larger-bodied and 'average' avatars. In this latter study, a bespoke exergame was developed that used the stimuli developed in the previous studies as avatars.

Results: The systematic review (Chapter 4) revealed that a small number of studies had investigated using avatars to promote health related behaviour, and a common finding was that using larger-bodied (compared with athletic) avatars in exergames resulted in reduced physical-activity. Since this was explained in terms of stereotypical behaviours, such as laziness, being assimilated into the players' behaviour, the qualitative study (Chapter 5) explored participants' accounts of being restricted to an athletic avatar and found that this was not always a positive experience. To explore this further, Chapter 6 investigated the stereotype structures of plus-sized and athletic bodies were both found to suggest ambivalence. Text descriptions of prototypical 'athletic' groups (Chapter 6), and images of virtual humans with athletic bodies (Chapter 7) were rated as

more competent, more arrogant, and less friendly compared with larger-bodied examples. Negative stereotypes, such as laziness, were reflected in evaluations of larger bodied representations, but so was the potentially positive trait of affability. In Chapter 8, larger-bodied exemplars that were animated with counter-stereotypical information (running on the spot) were rated less negatively than those that were stationary. When the exemplars were used as avatars, there was no evidence for behaviour change as a function of avatar-physique in the experimental replication study (Chapter 9).

Conclusion: By using methods derived from existing social psychological theories, it is possible to create representations of larger bodies that are evaluated more positively. Further, there may be negative consequences to relying on athletic-bodied avatars to encourage exercise. Although there was no strong evidence that participants behaviour was affected by the type of avatar used, an argument can be made for allowing users to explore a broader range of physiques and presenting larger-bodied characters positively as competent agents.

Contents

List of Tables

List of Figures

We fill pre-existing forms and when we fill them we change them and are changed.

Frank Bidart, *Borges and I*

David Foster Wallace, *The Pale King*

Chapter 1

Introduction

> Unpopular, overweight, and socially inept
>
> ---
>
> Kowert, Festl, and Quandt (2014)

The broad aim of this thesis was to address the question, *can the representation of the self in a videogame affect health-related behaviours?*. In 2011, the Internet Advertising Bureau estimated that there were 33 million gamers in the United Kingdom (IAB, 2011), and so videogames are a potentially fruitful medium for health persuasion. As previous authors have already noted, the avenues of persuasion in videogames (*rhetoric*) are singular, and may be utilised not just to influence, but also engage the audience in making value judgements (Bogost, 2007; Kolko, 1999).

According to Bogost, this level of interaction makes videogames a particularly persuasive medium, and both extols and warns against the use of *serious games* in a range of settings from health to environmental awareness. More specifically, Kolko (1999) argues that a particular element of videogame development, the design of the character a player controls (avatar), carries its own rhetoric and may be used to reinforce harmful stereotypes such as the portrayal of women as sexual objects. As will be discussed on page 52, stereotypical evaluations play a large role in what will be the major avatar persuasion theory of this thesis, the Proteus Effect (Yee et al., 2009). The most common manipulation of avatar appearance in the health-persuasion literature is weight, or body size. The current dominant message in many societies is that fit and strong are good, and fat and weak are bad. This manifests in either the absence of representation, or stigmatised portrayal of people with larger bodies, not exclusively in videogames, but in the media in general (Puhl, Peterson, DePierre, & Luedicke, 2013). This stigma is not limited to the videogame industry, but also extends to research on videogames, as will be discussed further in Chapter 4.

One of the objectives of this thesis is to challenge these value judgements, and to explore whether (and eventually how) the rhetoric of videogame design (specifically avatars) may be used to reduce negative stereotypical evaluations of digital characters, which in turn may have positive implications for in-game and post-game attitudes and behaviours.

Although stereotypical attributes of groups may be projected upon virtual *others* (Non-Playable Characters (NPCs), or agents), or the self in a virtual environment (avatars), the ultimate goal of the thesis is to explore how a representation of the self in a virtual environment may affect behaviours and beliefs. However, this groundwork could equally be applied to NPCs. Since the intended context is health, the aim of the thesis is to address health-related attitudes and behaviours, specifically exercise performance.

1.1 Exercise and Exergames

Much research has been conducted on the broad topic of *videogame health* (Rahmani & Boren, 2012). Such attempts include the development and implementation of so called *serious games*, in which the goal of the videogame goes beyond simple enjoyment (Wattanasoontorn, Boada, Garcia, & Sbert, 2013). Serious games may present a message, or deliver training through the gaming narrative or tasks (e.g. Darville et al., 2018); involve networked 3D virtual environments through which health communication is delivered, and activities such as exercise and food shopping are practised via an avatar (e.g. Behm-Morawitz, Lewallen, & Choi, 2016); or directly involve exercise, in so-called *exergames*, which aim to reduce the amount of sedentary time, and promote physical activity the home (Lwin & Malik, 2012). The latter suggestion of using videogames to promote physical activity (*exergamification*), generally involves the use of motion detecting peripherals to allow body movements as the control method for videogames (Matallaoui, Koivisto, Hamari, & Zarnekow, 2017)[1].

The effectiveness of these exergames has been well documented in both the Health Psychology and Human-Computer Interaction literature. These types of game have been demonstrated to be as effective as light exercise, but with the benefit of being performed in the privacy of one's own home and have been applied to outcomes such as reducing sedentary time, weight control, and encouraging physical activity (Lwin & Malik, 2012; Matallaoui et al., 2017; Peng, Lin, & Crouse, 2011). There is no research on whether exergaming was ever successful as a population level intervention. Further, the commercial exergame bubble appears to have burst, with the discontinuation of the Microsoft Kinect device, Nintendo focussing on portable gaming, and Sony investing in Virtual Reality. As Matallaoui et al. (2017) suggest, this reduction in consumer interest may be due to the *Novelty Effect*, in which the excitement surrounding a product dissipates over times and is not recharged.

If exergaming is potentially obsolete, then why conduct a research programme that focuses on promoting exercise with videogames? More attention is being paid to emerging health risks associated with sedentary lifestyles (Vainshelboim, Brennan, LoRusso, Fitzgerald, & Wisniewski, 2019). These appear to be independent of levels of physical activity, and so physical activity is

[1] This could cynically be seen as the gaming equivalent of a tricking children into eating vegetables by mixing them in with a dessert.

not necessarily protective against sedentary behaviour, if guidelines are developed and widely publicised, then a resurgence in the popularity of active videogaming may occur. It is also interesting to note that during the global COVID-19 pandemic that began in late 2019, Nintendo reported significant shortages of the exergame RingFit as social distancing and lockdown conditions became more diffuse, suggesting that exergames may be seen as a last resort for physical fitness in times of crisis (D'Anastasio, 2020).

A key feature of many videogames (including exergames) is the inclusion of a videogame *avatar*, or a representation of the self in a virtual environment (Nowak & Fox, 2018). Traditionally these are seen from a third person perspective. However, with developments in VR, immersive embodiment (e.g. The *body swap* illusion) is becoming more mainstream, and so interest in the reflexive effects of avatar appearance on the self in VR may re-emerge. Finally, exergames require direct behaviour as an input, and unlike abstract concepts such as *assertiveness* or *prosociaility*, may be directly measured (e.g. through in-game performance, number of movements made, etc), and require on fewer (if any) ostentatious ruses involving confederates than previous studies on avatar-behaviour assimilation[2].

In many games, players may customise an avatar to look like themselves. These games provide a number of *affordances* or options that the participant may change. In some cases, default avatars are provided, or a selection may be made from a series of generic or pre-built avatars. These avatars often have athletic, or idealised physiques which reflect the replacement of *thinness* with body tone and muscularity as a dominant Western ideal (Schaefer et al., 2015). There has been some work on how these idealised forms in both real and virtual models may affect a person's internal perception of their body (their *body image*), which will be outlined in Chapters 2, 4, and 5.

[2]Some of the lengths that the early VR pioneers had to go to were admirable, see discussion in Chapter 2, from page 26.

1.2 Objectives

The objectives of this thesis are as follows:

1. Assess the strength and quality of evidence for avatar appearance as a health-related behavioural intervention.

2. Explore the experience of playing an exercise game with a facially similar, athletic avatar.

3. Determine the stereotype space that muscular and larger bodied avatars occupy, compared with larger bodied avatars.

4. Determine whether activation of these stereotypes affects in-game behaviour.

In summary, the programme has two key aims: the first is to assess the state of theory and implementation of research into the health-behavioural modification properties of avatars, and to contrast this with the reported playing experience of gamers playing an exercise videogame with athletic avatars (Chapters 2, 4, and 5). The second is to critique previous research (e.g. Li, Lwin, & Jung, 2014) by deconstructing avatar persuasion theories and attempting to modify how avatars are evaluated (see page 142).

A series of studies was designed in Chapters 6 - 8 to address objectives 2 and 3. They assessed how athletic and plus-sized bodies are internally represented through evaluations of verbal descriptors and 3D synthetic humans (*exemplars*). Such evaluations were expected to be heavily influenced by stereotypes, which is a necessary condition for direct avatar persuasion (see pages 57 and 230). Next, attempts were made to interrupt these stereotypical evaluations using counter-stereotypical information (Chapter 8). Finally, whether these evaluations may be manipulated and translate into behaviours within an exergame, as predicted by the Proteus Effect was investigated in Chapter 9.

1.3 Structure of the Thesis

The structure of the thesis is described in the following paragraphs:

Chapter 1

In this chapter[3], the structure of the thesis is being introduced, the aims and objectives were formally stated, and a chapter-by-chapter summary is currently underway.

Chapter 2

In Chapter 2 the relevant theories and supporting evidence will be reviewed. This will begin with theories of avatar embodiment and how this relates to self-perception. Next, theories of avatar influence will be reviewed, covering the main previous contribution to this area: the Proteus Effect (Yee et al., 2009).

Chapter 3

The methodology that was used in the thesis will be reviewed. This will include sections on the use of mixed methodology, which is the approach that will be employed in this thesis; the population of interest and sampling plan; and the adoption of Open Science principles. There will also be a discussion of the philosophies of science that are adopted throughout the thesis, the main methods employed in the empirical chapters (systematic reviews, interviews, experiments, and replications), and an argument will be made for the use of Bayesian statistics as the statistical paradigm with which to model the data.

[3] A chapter summary is complete iff it contains the chapter it is contained within.

Chapter 4

In the first research chapter of the thesis, findings from a systematic review of studies investigating the effect of avatar-appearance on health-related attitudes, beliefs, and behaviours will be reported. This will combine the elements discussed in Chapter 2 into a coherent picture of the current state of the theory and methods used in the field of avatar health persuasion.

Chapter 5

The first empirical study of the thesis will detail the findings from a qualitative interview study. Participants were interviewed about their gaming and avatar generation habits, body image, and health-related behaviours prior to playing an exergame with an idealised *champion* avatar. The data were analysed using inductive Thematic Analysis (Braun & Clarke, 2006), and resulting themes were interpreted reflexively using elements from Critical Realism (Bhaskar, 2016).

Chapter 6

In the first quantitative study, a group of participants were asked to generate words that they felt were stereotypically associated with larger bodied, athletic-looking, or fit people. The most frequently occurring words were rated by a second group of participants for applicability to groups of people with *athletic* or *plus sized* body types. The set of stereotypically applicable words is presented, as well as a preliminary factor structure. Differences in applicability ratings between verbal group descriptors are explored using multi-level Bayesian regression analysis.

Chapter 7

In Chapter 7, a group of participants rated the applicability of the words from Chapter 6 in response to synthetic virtual humanoid exemplars of larger bodied and athletic individuals. The factor structure for the ratings of these words will be assessed and compared with the data from the verbal descriptors in Chapter 6. The findings will be discussed from an ontological perspective, in which an exemplar inherits attributes from its reference class.

Chapter 8

In Chapter 8, the same set of words used in Chapters 6 and 7 were rated for applicability in response to animated versions of the exemplars from Chapter 7. The animations provided *counter-stereotypical* information to the larger-bodied exemplars (i.e. running on the spot) to determine whether negative evaluations of the exemplars may be improved with additional information.

Chapter 9

In the final empirical study, the animated exemplars from Chapter 8 were used by participants as avatars in a bespoke exergame. The method broadly replicated a previous study (Li et al., 2014), in which participants moved slower during an exergame when using a larger-bodied avatar. In this study however, the avatars had been validated for stereotype content, and the potential ambivalent stereotype content of athletic avatars on perceived performance was investigated.

Chapter 10

In the concluding chapter, the findings as a whole will be synthesised and discussed in the context of the existing literature. Next, the research process will be reflected upon, as well as a reflexive discussion of The Author's position in the research process. The findings will then be contextualised, followed by a discussion of the strengths, weaknesses, and future directions.

Chapter 2

Literature Review

This chapter will lay the theoretical foundations of the thesis. First, the target behaviour (physical activity) will be defined, and some of the contributions made by health psychology to promoting this behaviour will be discussed. Next, the target population will be described. Following this, *exergamfication* will be introduced as a potential intervention that may promote physical activity in the target population. The contribution of avatars, particularly their appearance, to in-game and post-game behaviours will be discussed using the Proteus Effect (Yee et al., 2009) as a theoretical framework. The link between avatars and stereotypical evaluations will then be made, followed by a review of stereotypes related to weight stigma, fat phobia, and "dumb college jock" stereotypes, and their relations to plus-sized and athletic physiques respectively. Following this, the research question will be framed as a study into the representation of body types in exercise-based videogames, and whether they may be used to reduce negative evaluations of larger bodies, and by extension increase physical activity.

2.1 Physical Activity

Definitions

Caspersen, Powell, & Christenson (1985) propose the following succinct definition of physical activity:

> "any bodily movement produced by skeletal muscles that results in energy expenditure".

The health benefits of physical activity are well documented and cover many types of activity. As will be demonstrated, physical activity is associated with the decreased risk of a variety of preventable diseases such as heart disease, metabolic syndrome, diabetes, and stroke ("Benefits of exercise," 2018).

Examples of physical activity include occupational, sporting, conditioning, or household activities, and are typically measured in total caloric expenditure[1], or more recently through Metabolic Equivalent Tasks (MET), a measure of task vigour which is proportional to the caloric expenditure per kilogram per minute (Jette, Sidney, & Blümchen, 1990). Standard definitions of vigour are 1 MET being stationary and seated, 1-3 being light intensity physical activity, 3-6 being moderate, and greater than 6 being vigorous. Exercise is a special case of physical activity that is *planned, structured, and repetitive* (Caspersen et al. (1985), p126) with the specific goal of gaining or maintaining physical fitness. Finally, Caspersen et al. (1985) define physical fitness as a set of abilities that are related to health (e.g. endurance, strength) or skills (e.g. balance, agility) which may be assessed using a relevant task.

Physical Activity and Health

In a systematic review and meta-analysis of risk protective studies, Samitz, Egger, & Zwahlen (2011) identified 6 domains of physical activity: Leisure time physical

[1] A calorie is the amount of energy required to bring 1mL of water to $100^{\circ c}$

activity, which involves any recreational activity involving physical activity such as hiking, golf, swimming, sports, etc; exercise and sports, including structured activities involving aerobic or muscle-strengthening; physical activities of daily living, which includes house work, gardening, walking, climbing stairs; physical activity for transportation, including cycling or walking to work; and occupational physical activity such as bar work, building, etc. They report that the risk of all-cause mortality may be reduced by an increase in physical activity, but that some domains are more effective than others. This association was stronger for women than men, and the risk was moderated by the domain, such that structured, and exercise activities were more protective than occupational and transport activities.

Evidence has been presented suggesting that physical activity may reduce the risk of conditions such as cardiovascular disease (Berlin & Colditz, 1990; Nocon et al., 2008), stroke (Lee, Folsom, & Blair, 2003), colon cancer (Wolin, Yan, Colditz, & Lee, 2009), cognitive decline (Sofi et al., 2011), as well as benefitting mental health (Craft & Landers, 1998). Further, the state of *physical fitness* has also been linked with lower all-cause mortality (Blair et al., 1989). The benefits appear to have an inverse non-linear relationship, with the largest benefits occurring between no-activity and low-levels of activity, however the risk is consistently reduced with more activity but at a smaller rate (Woodcock, Franco, Orsini, & Roberts, 2010).

The evidence of benefits for physical activity is strong enough for guidelines to be produced by a variety of health organisations including the world health organisation (World Health Organisation, n.d.). Specific guidelines differ, for example the guidelines from the medical officer's report states:

> For general health benefit, adults should achieve a total of at least 30 minutes a day of at least moderate-intensity physical activity on 5 or more days of the week. The activity can be lifestyle activity or structured exercise or sport, or a combination of these.

Whereas in the USA, guidelines are more varied (United States. Department of Health, 2008):

> Adults should do at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes of vigorous-intensity aerobic physical activity. For additional and more extensive health benefits, adults should increase their aerobic physical activity to 300 minutes a week of moderate-intensity or 150 minutes a week of vigorous-intensity aerobic physical activity.

Although the duration and intensity may differ, the consensus appears to be the equivalent of 150 minutes of moderate intensity exercise per week as a minimum (Samitz et al., 2011).

Psychology and Physical Activity

Since there is a consensus on the myriad benefits of physical activity in daily life, encouraging inactive people to be more active may be seen as a form of *preventative treatment*. Since factors like beliefs, attitudes, and perceived barriers may influence whether a person will engage in physical activity, a great deal of research has been conducted by health psychologists into ways to encourage physical activity. Moreover, the psychological barriers to physical activity have been studied using both qualitative and quantitative approaches, and theoretical models aiming to explain the performance of certain behaviours have been developed.

Barriers and Benefits of Exercise One factor that may prevent someone from participating in a particular activity is the perception of barriers. Research suggests a wide range of potential barriers, with some common themes emerging. For instance, Myers & Roth (1997) report a four-factor model of perceived barriers in a sample of undergraduates from Alabama, USA. The first, Time and Effort, includes the convenience of performing exercise, the predicted enjoyment of the activity, judgements about the self being too lazy are contributors. Physical barriers include the perception that exercise will look silly, that it will lead to discomfort, and that the person is too fatigued to exercise. Social barriers include not having social support, and not being encouraged. A final category of barriers refers to specific instances such as weather, medical problems, and work and social priorities. In a cohort of older Finnish participants, concern about injury and discomfort was a widely held belief (Rasinaho, Hirvensalo, Leinonen, Lintunen, & Rantanen, 2007). In a sample of participants from the general population of Australia, Ross & Melzer (2016) identified a broader range of barriers, including expense, failed previous attempts, fear of injury, intimidation, and access to facilities. They found that these factors were associated with reduced participation in physical activities. More recently, in a large scale prospective survey study, More, Phillips, & Colman (2019) found that although body dissatisfaction did not directly affect duration of exercise, there was a mediated effect between dissatisfaction, avoidance, and exercise duration, which itself was mediated by embarrassment and fatigue, leading the authors to suggest that these two barriers are important predictors of exercise.

Positive beliefs about exercise also predict the engagement with, and maintenance of physical activity. Brown (2005) found that perceived benefits were better predictors of physical activity than perceived barriers. Myers & Roth (1997) estimated the perceived benefits of exercise in their Alabama cohort, and found that improved body image, good mood, better health, and sociability were key factors. They also found that strength of belief in the perceived barriers in all four factors was reduced depending on participants stage of change in regard to physical activity. This is a reference to the Transtheoretical Model of behaviour change whereby a person moves through a series of phases representing the gradual increase in likelihood that a behaviour will be adopted and maintained (Prochaska & DiClemente, 1983)[2]. Perceived benefits have also been positively

[2] A further health behaviour change model is the Theory of Planned Behaviour (Ajzen, 1991),

correlated with an increase in self-reported regular exercise habits, whereas barriers were correlated with short term, sporadic attempts (Grubbs & Carter, 2002). As will be suggested below in section 2.3, videogames may offer a solution to these perceived barriers, but first the potential population of interest will be described.

2.2 Videogames and Gamers

Videogames remain strongly associated with *nerd culture*[3] , and as such stereotypical associations of gamers being overweight, socially inept, and mostly male abound (Kowert, Griffiths, & Oldmeadow, 2012). However, as will be discussed later (page 46), fitness and body-size have been demonstrated to be orthogonal. Moreover, as shown in section 2.1, additional exercise at most levels of fitness is beneficial, with the slope flattening out towards higher levels of fitness. If one assumes that gaming habits and body size are unrelated, then all but participants at the extreme tail of the fitness distribution would benefit from replacing some gaming time with additional exercise. This brings us to consider the population of people who play videogames.

Who Plays Videogames?

Videogaming is identified as a sedentary activity, in line with television watching (Garn, Baker, Beasley, & Solmon, 2012), which may explain the *overweight* element of the gamer stereotype. If this stereotypical perception represented the *true state of affairs*, then this project would best be targeted at people who are gamers, who would automatically be male and overweight. However, a cursory look at the global marketing research on gaming and gamer demographics suggests that this is not the true state of affairs.

which is briefly discussed on page 18.

[3] Pop culture that is heavily rooted in technology, Kelly (1998).

In attempting to influence people's behaviour using videogames of any type, reflection on the population of people who play games is important. Without this information, an intervention may end up not being tailored to the most receptive demographic. The Interactive Software Federation of Europe (ISFE) provides information about the consumption and use of videogaming products. Through a project called GameTrack, each fiscal quarter a sample of the population from European countries is asked about their gaming preferences. In their 2019 report (covering data from 2018), they reported that 74% of 15-24-year olds, 67% of 25-34-year olds, and 49% of 35-44-year olds play videogames, with the 25-34-year olds being the fastest growing group. Of all gamers between the ages of 16-77, 77% play games for at least an hour a week. The report also shows that, in the EU, 47% of women are gamers, which heavily contradicts the male gamer stereotype. Although Europe is a heterogeneous set of countries, these numbers also broadly reflect the UK by itself, with 56% of 16-69 year olds reporting playing videogames "on most days", and 50% of gamers being women (Ukiepedia, 2020). Moreover, both the USA ("2019 essential facts about the computer and video game industry - entertainment software association," n.d.) and China ("The chinese gamer," n.d.) report similar proportions of gamers across gender and age ranges. Of note is the absence of gaming demographics from African countries. The most information I could find was that the South African gaming market was increasing (Statistica, n.d.), and that Egypt, South Africa, Morocco, Nigeria, and Algeria were the largest consumers of videogames.

With such a heterogenous population of people who play games, it is unlikely that the stereotype described above, and in the Chapter 1 epigraph, is accurate. In summary, given the representation of University aged, or recently graduated individuals in the gamer demographic surveys (ages 16-44), sampling of University students for the gaming elements of the thesis is arguably acceptable[4]. The population and sampling assumptions made in the thesis are discussed in-depth in Chapter 3, from page 71.

[4] If not entirely ideal since there may be a restriction of range across some variables.

What do they play?

The games that people enjoy may be inferred from various sources, such as usage and sales data. Some of the most popular games globally are multiplayer death match and team deathmatch games such as Crossfire and Dungeon Fighter. Naturally, data are only available for online games, since there is no way to monitor offline play.[5]

Although the highest selling games of all time are Minecraft ("Minecraft's player count continues its upward trajectory," n.d.) and Grand Theft Auto V (Makuch, n.d.), these are both cross platform titles, meaning that they are available on different consoles and PCs. Interestingly, the highest selling console videogame of all time is Wii Sports, an active videogame, which is the type of game used by several researchers to influence health-related behaviours (see section 2.3). Globally, Wii Sports[6] sold nearly 83 million units (Nintendo, 2020b). This number needs to be interpreted cautiously, since Wii Sports was bundled with the Nintendo Wii console, this means that it was necessarily purchased by everyone buying a new console. Perhaps more telling of the popularity of exercise videogames at the time was the popularity of Wii Sports Resort (33 million copies sold), and Wii Fit and Wii Fit Plus (nearly 44 million units sold). As mentioned in Chapter 1, exergaming potentially existed in a bubble, however Nintendo have reported that RingFit Adventure on the current generation console sold over 2 million copies worldwide, and the rhythm boxing game Fitness Boxing has sold upwards of half a million copies, suggesting that there is still a market for exergaming(Nintendo, 2020a).

[5]Deathmatches are a mode of play in which players, either in teams or independently, attempt to kill one another and scoring *Frags* (units of death).

[6]A selection of sports games that are played using a motion sensitive controller.

2.3 Exergames

Attempts have been made to append physical activity to traditionally sedentary behaviours. The most prominent example is *exergamification*[7], which involves the use of full body movements to control a game[8]. These games often use additional peripherals that detect motion so full body exercise can be performed to control a game. These products have the two-fold benefit of reducing sedentary time and increasing physical activity. Exergaming also has the benefits of requiring a potentially smaller financial investment (if one already owns a console), may be played in the privacy of one's own home, and are generally self-contained so extra facilities are not required - meaning that they may circumvent many of the barriers discussed earlier (12).

The benefits of exergaming are well reported. They have been shown to be as effective, in terms of energy expenditure, as more traditional forms of exercise (Matallaoui et al., 2017; Peng, Crouse, & Lin, 2012; Peng et al., 2011; Sween et al., 2014), and may have the potential to increase physical fitness in the long term (Huang, Wong, Lu, Huang, & Teng, 2017). There are several varieties of exergame, six of which were studied by Bailey & McInnis (2011) in terms of energy expenditure and enjoyment. These included Dance Dance Revolution in which a player uses their legs to activate pressure points on a play mat in time to music; Wii Sports Boxing and Xavix Boxing, in which players mimic punching movements to hit an opponent; and a variety of commercial exercise games involving wall-mounted pressure sensitive pads (LightSpace, SportsWall), and camera-based motion detection (Trazer). Average METs for each of the games were above resting and ranged between 4-8, with the Wii being the lowest and most comparable to treadmill walking at 3MPH. Further, it has been demonstrated that exergames may be associated with greater uptake and maintenance of vigorous physical activity in a clinical trial (Bock et al., 2019).

[7] A less current example would be TV-based fitness shows such as Mr Motivator .

[8] Not necessarily videogames - there are some apps that gamify running by creating goals surrounding a zombie apocalypse that the user must complete whilst avoiding attacks.

Psychology and Exergames

In addition to simply increasing energy expenditure, research has been conducted on the extent to which exergaming can affect Psychological constructs such as self-efficacy, perceived barriers, and behavioural intentions. Lyons & Hatkevich (2013) assessed the extent to which a set of 18 fitness videogames implemented evidence-based behavioural change strategies. They found that modelling behaviours, specific performance feedback, reinforcement, energy expenditure feedback, and guided practice were the most common methods, with action planning being the least prominent. The majority of games also had social functionality. The authors report three categories of behaviour change techniques, including those that enhance self-efficacy, those that target self-regulation and habit-forming, and 'other' types which involve rewards and social functionality, such as multiplayer modes.

There has been some work on predicting changes in constructs such as behavioural intentions, attitudes, and perceived behavioural control from of the Theory of Planned Behaviour (Ajzen, 1991; Fishbein & Ajzen, 1975) . In a cohort of children (N = 1112) between the ages of 10 and 12, exergaming was associated with improvements in attitudes towards exercise, perceived behavioural control, and subjective norms (Lwin & Malik, 2012). In young adults, similar but more complicated associations have been found. Van Nguyen et al. (2016) claim that in a randomised controlled trial, attitudes and intentions towards exercise were strengthened in people who frequently exercised but reduced in those who did not. However, this claim is supported by very weak evidence. Six reported p-values were reported as significant, but only one value was below the assumed alpha threshold of 0.05[9]; moreover, no effect sizes confidence intervals are provided in the paper, and there is a seemingly arbitrary separation of frequent and infrequent exercisers based on a cut point of twice a week. Considering that the main message of the paper is reliant on this categorisation, more justification, and ideally a

[9] Assumed because it is not explicitly stated.

pre-registered replication would be needed[10].

Intentions to exercise in the future were improved in participants in a repeated measures study that compared *traditional exercise* with games available for the consumer product *Wii Fit* (Garn et al., 2012). Intentions were higher after playing Wii Fit than traditional exercise in participants across three BMI ranges ("healthy-obese"). Intentions to play were also indirectly associated with duration of play, mediated through self-reported enjoyment, feelings of competence, and the feeling that younger adults were present in game (Limperos & Schmierbach, 2016). Intentions are a useful short-term construct for predicting future behaviour (Armitage & Conner, 2001; McEachan, Conner, Taylor, & Lawton, 2011; Webb & Sheeran, n.d.), and intentions to play exergames have been shown to correlate with intentions to perform other exercise in children (Li & Lwin, 2016). Two further mediators in the Li & Lwin (2016) study were self-presence and avatar identification. The use of avatars as modifiers of behaviour is the primary focus of this PhD project, and the upcoming sections will assess research suggesting that they may effective in promoting exercise behaviour through exergames, and the potential situations where they may interfere.

In summary, exergames offer a unique and promising avenue towards the promotion of physical activity. As this thesis develops through Chapters 4 and 5, the focus will narrow down to improving the intensity of exercise *when a person is already playing an exergame.* The mediator between the player and the exergame is often a virtual representation of the self. The potential of these representations to affect exergame outcomes will be discussed in the following sections.

[10] A further issues with this study include the possibility that the paper is salami sliced (Chambers, 2017), i.e. using the same dataset to produce many different papers and treating them as entirely different and independent studies (see Huang et al; 2017).

2.4 Avatars

Definition

The word *avatar* is taken from *avatāra*, the Sanskrit word for *descent*,[11] which refers to the descent of a deity into an earthly form in Hindu theology (Messinger et al., 2008). In computer-mediated communication, the word is rather more profane, and was popularised in the science fiction novel *Snow Crash* by Neal Stephenson where it referred to the representation of the self in the *Metaverse*, a fictional, virtual reality version of the internet (Stephenson, 2014)[12].

There is a great deal of variation in how avatars as digital entities are defined, with some authors arguing that they are strictly visual representations (Nakamura, 2013), or necessarily humanoid (Messinger et al., 2008). Nowak & Fox (2018) argue that this definition is underspecified, excluding experiences of text-based role-playing game players, or people with visual impairments. They note that a common thread is the importance of avatars for representation, which was almost ubiquitous in their readings (Nowak & Fox (2018), p31). Instead, they propose a broader definition of *avatars* that encompasses all modalities, and *future proofs* against developments in interfaces:

> [...] we endorse a more open definition and argue that an *avatar* is a digital representation of a human user that facilitates interaction with other users, entities, or the environment. [p34]

This definition encapsulates all virtual experiences in which mediated interaction is necessary. For example, avatars are not needed for general internet browsing, but are when navigating a virtual environment or communicating with others in a chatroom. Avatars are not necessarily visual 3D models and may also

[11] अवतार.

[12] The first use of the word in a videogame was in the 1985 game Ultima IV, a game which I spent many nauseating hours trying to play on the Sega Master System with my brother in the 1990's

represent a person in the form of text, sound, haptics or through a mixture of several modalities (e.g. a World of Warcraft[13] player may have a 3D avatar and a screen tag/username).

Avatar uses and variation

Nowak & Fox (2018) discuss several dimensions on which that an avatar may vary. For instance, avatars may vary in similarity to the player (homophily), the extent to which the avatar appears human (anthropomorphism), and realism (both behavioural and visual). A common feature of videogames is avatar customisation, and often the user is given options to create an avatar in their own (or an *others*) image (see page 144). As will be discussed in Chapter 5, the importance of these dimensions differ between people, and may vary based on how players relate to their avatars. These relations may serve as: a method of identity expression, exploration, and deception; mere functional purposes such as tools or extensions to the physical self; or a means to an end, i.e. the *avatar as an object* relation[14] (Banks, 2007, p. p185).

In closing their paper, Nowak & Fox (2018) discuss the future of avatar research, including a call to provide more methodological detail about the avatars used in research; researching the effect of embodiment duration on behaviour; and potential ambivalent effects of avatar use[15]. This suggestion informed the second dependent variable in Chapter 9 which attempted to measure *Arrogance* induced through the use of an athletic avatar.

Although there are fascinating questions to be asked in exploring identification with avatars of other modalities, in this thesis avatars will be operationalised as 3D humanoid (i.e. anthropomorphic) models. Moreover, implicit in the definition of Nowak & Fox (2018) is that an avatar is only an

[13] A massive multiplayer role-playing game.

[14] Akin to a counter or game piece

[15] For example, as will be discussed in section 2.5 an attractive avatar may lead to more confidence, but also may lead to greater narcissism or self-objectification.

avatar when it is controlled by a human; as such the distinction between avatars and *exemplars* will be made. In Chapters 7 and 8 participants did not control the character, but merely observed and evaluated it. As such, the characters *exemplify* the characteristics of a group, without being controlled by machine or person[16], whereas in Chapters 5 and 9, the characters were (somewhat) under the users control, and therefore qualify as avatars.

Player-Avatar Relations

According to Banks (2007), avatars may serve different purposes depending on the perceived relationship between the avatar and the player. In a study that was based upon 70 hours of unstructured interviews in World of Warcraft, Banks developed a taxonomy of Players-Avatar Relations (PAR). This taxonomy consists of four categories: *Avatar as Object* in which the avatar is merely a tool, or game-counter which playing a game necessitates the use of; *Avatar as symbiote* in which the avatar is regarded as a separate but dependent part of the self, and so players feels that they must meet the avatars needs; avatar as me in which the avatar is regarded as the self; and the avatar as other, in which the avatar is regarded as a friend or colleague but not as the self. Each type of relation corresponds to a combination of the variables: intimacy, self-differentiation, and agency (figure 2.1). For example, the Avatar as object category is formed of low intimacy, high self-differentiation, and high player agency.

In further refining the concept of Player-Avatar Relations, Banks and Bowman (2016) developed a variety of measures to capture Player-Avatar Interactions (abbreviated PAX), most recently the Common Player Avatar Interactions (cPAX). According to Banks, Bowman, Lin, Pietschmann, & Wasserman (2019), PAX will vary depending on the PAR that a player has with their character. The cPAX measure conceptualises PAX in terms of four dimensions, Anthropomorphic Autonomy (the extent to which an avatar can

[16]This framing is also used by Ahn (2016).

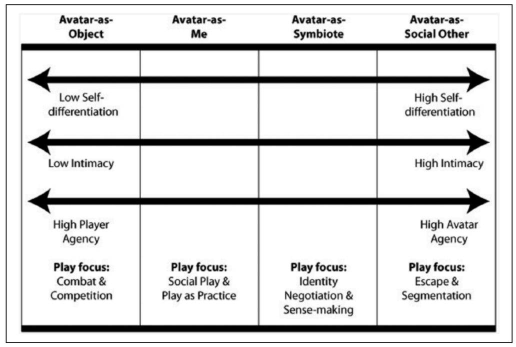


Figure 2.1: The taxonomy of avatar relations presented by Banks and Bowman (2016)

make its own choices), Relational Closeness (the level of emotional involvement a player has with the avatar), Critical Concern (the degree to which story inconsistencies can be ignored), and sense of control (Banks et al., 2019). Banks et al. (2019) found that the cPAX dimensions were predictive of Player-Avatar Relation style (as measured using a free-text and heuristic categorisation method): for instance, players who had *as me* or *as symbiote* relations scored higher on the Relational Closeness subscale. Elsewhere, PAR-PAX have been found to be predictive of the feeling of spatial and self-presence (Luo, Westerman, & Banks, 2019), which are discussed in more detail below (section 2.4). For example, Relational Closeness was associated with higher spatial-presence; the more aligned with the self the avatar was, the more mastery over the virtual environment players felt. Additionally, Relational Closeness and Sense of Control of an avatar predicted Self-Presence; a greater feeling of control and alignment with the self was associated with a stronger sense that it was *the self* who was in the game (see page 25).

It is important to note that the PAR-PAX constructs are based on data from MMORPG players discussing and referring avatars with which they may have spent hundreds, if not thousands of hours in game. It is possible that relations formed over such a long time are qualitatively and quantitatively different from those formed on an *ad hoc* basis during experiments, or casual play. The point here is although PAR-PAX offer a useful description of long-term avatar relations, it is as of yet unknown if these represent a *style* of interaction that generalises across multiple avatar interactions, nor is there an extant tool that measures this.

Presence

Arguably a central function of an avatar (depending on the purpose intended by the user) is to promote the illusion of *presence*, which in the taxonomy proposed by Lee (2004) is defined as:

> "a psychological state in which virtual [objects/actors/selves] are experienced as actual [objects/actors/self] in either sensory or nonsensory ways".

The variables in the brackets relate to the three aspects of Lee's presence typology: physical, social, and self-presence. Physical presence relates to the virtualisation of physical experiences and can include invoking a sense of place as well as interaction with objects and the environment. In previous versions of physical presence, the definition included the sense that one has been *teleported* to a new environment, which Lee's physical presence does not require. This was important because previous work had been conducted in high-fidelity immersive virtual environments, where *transportation* was the aim, but presence may be experienced via substantially less immersive environments - such as televisions - through which the viewer could not sensibly claim to have been transported, but could nevertheless feel that they are present in the depicted environment.

Social presence may be discussed in terms of the work on *para-authentic* social relationships . Indeed, early theories describe the notion of *parasocial* interactions, in which an audience member forms a one-sided relationship with a media personality (Horton & Richard Wohl, 1956; Klimmt, Hartmann, & Schramm, 2006). For example, in figure 2.2, a dialogue between a former adventurer and the *Dragonborn* protagonist of Skyrim is presented. The characters refer to the avatar using a second person pronoun which may refer to the avatar, but also be directed through the fourth wall at the player. Indeed, in Lee's definition of social presence, the intelligence of a virtual agent is simulated, and the artificiality of the interaction is un-noticed by the user.

Figure 2.2: An example of a parasocial interaction in the game Skyrim

Perhaps most relevant to the current programme is the notion of self-presence, which is achieved when the virtuality of the self in a virtual environment ceases to be perceived. This was further unpacked by Ratan (2010), and itself has three levels - broadly based on (Damasio, 1999) elements of the self - which are: *Proto, Core, and Extended* self-presence. The proto-self is the physical representation of the self in a virtual environment and the integration of the representation into the user's body schema. The core self refers to emotional interactions with the content of the virtual environment. High core-self presence means that the user is emotionally engaged with the scenario and its actors and experiences these emotions vicariously. The extended self represents the ways in which the virtual representation projects some part of the user's actual identity. Ratan suggests that the extended self-presence may be moderated as a function of identification with a self-representation (avatar), or the time taken to create this representation. That is, self-presence is dependent on there being a self-representation in a virtual environment.

2.5 Avatars, Behaviour Change, and the Proteus Effect

Much work has been conducted on the use of avatars as a tool for modifying behaviour. This assumption is the foundation of the current thesis. There are a number of existing theories that claim to predict the influence of avatars. Two major contributions were PhD projects produced by students from the Stanford Human Technology Interaction laboratory in the late 2000's. The Proteus Effect (Yee, 2007) focused on behavioural assimilation through avatar embodiment; and Virtual Self Modelling (Fox, 2010) focused on increasing exercise behaviours using self-similar virtual representations. The focus of this thesis is on the Proteus Effect [17].

The Proteus Effect

The Proteus Effect is named after the Greek god, Proteus (figure 2.3), who could change the form in which he appeared to humans (Yee, 2014). The essential idea of the Proteus Effect is that by changing their perceived physical appearance, people will make inferences about themselves, and adapt their attitudes and behaviours to fit with the new form. To quote Yee et al. (2009):

> …in line with self-perception theory, they conform to the behaviour that they believe others would expect them to have. [p274]

This heavy grounding of the Proteus Effect in Daryl Bem's Self-Perception Theory warrants a brief discussion of the assumptions and predictions made by the theory.

[17] The Author is however running a replication of the VSM studies which initially were intended for the thesis - however due to time constraints and a change in direction, the study became a *side hustle* [18]

[18] Phrase taken from Maddi Pownall's talk at PsyPag 2019.

Figure 2.3: Would it even be a thesis without some form of ancient Greek Philosophy? Image of the Greek God Proteus taken from Alciati's Emblematum Liber

Self Perception Theory

The Proteus Effect is based largely on Self Perception Theory (SPT) which predicts that attitudes may be formed or modified as result of observing one's own behaviour from a third person perspective (Bem, 1972); in other words, behaviour paradoxically precedes attitudes. In a classic SPT study, two groups of participants will perform the same written task under two conditions, e.g. receiving a large or reward for a writing task that involves arguing against a topic they feel strongly about (e.g. the dangers of free speech on campus). The findings suggested that the attitudes of people who received a smaller reward shifted away from their pre-essay state towards agreeing with the essay. Bem suggests that because the reward was small, these participants believed they had written the essay of their own volition, made an inference about themselves based on their behaviour, and adapted their attitudes to conform with this behaviour. SPT has also been demonstrated in varying conditions of perceived free-will (Bem & McConnell, 1970), and when participants were forced into making false confession (Bem, 1966).

Elsewhere attitudes towards individuals have been manipulated by manipulating participants interoception (perception of physiological activity). Valins (1966) made male participants listen to a frequency generator that they believed was their heart rate. They found that when an increase in frequency was paired with pornographic images, participants rated the model as more attractive and was more likely to select the pornographic image as a 'reward'. The former finding has received some support through replication, although with caveats regarding the role of attention (Stern, Botto, & Herrick, 1972). Further, Dutton & Aron (1974) found that male tourists who were approached by a female researcher in a high-arousal context (perilous-seeming bridge) were more likely to contact the researcher than those in a low arousal context). Ethical issues and heteronormative assumptions aside (these were conducted in a different era), the conclusion in both of these studies was that participants attributed the arousal (perceived or actual) to finding the researcher attractive.

Enclothed Cognition

The symbolic content of appearance is also a feature of the Proteus Effect. In order for inferences to be drawn from an altered version of the self, there must be pre-existing attitudes towards a stimulus. The symbolic content of apparel, and how this may affect behaviour (and the perception of behaviour), was investigated in a series of studies by Frank & Gilovich (1988). Using the penalty records of National Football and National Hockey League (NFL and NHL) teams, the authors found that players wearing black uniforms were more aggressive than those with other colours (Frank & Gilovich, 1988, Study 2). Expanding on this finding, Frank & Gilovich (1988) investigated whether self- and social-perception could be altered by manipulating clothing colour. In one experiment (Study 1), participants rated the aggression of a defensive team in one of two videos of sets of American football plays. The defensive team either wore a black or a white uniform, whilst the offensive team wore a red uniform. A group of 20 experienced referees rated the black uniformed team as more aggressive than the white uniformed team[19]. In a final experiment (Frank & Gilovich, 1988, Study 4), people wearing either black or white uniforms selected more aggressive activities to play than when they were wearing their own clothes, or than those wearing white uniforms.

The authors of these studies claim that the colour black has semantic connotations of evil and death, and as such are associated with aggression when seen and worn. The semiotics of clothing were further investigated by Adam & Galinsky (2012), who investigated changes in performance on an attentional task when the same garment was assigned different symbolic meaning. Participants performed a visual attention task whilst wearing a lab coat described either as a doctor's coat, or a painter's coat. To control for simple visual priming, participants in a third condition performed the task after having seen a doctor's coat displayed on a table. Participants demonstrated higher attention, as measured by accuracy on a Stroop task, when wearing the doctors coat than those in the other two conditions (Adam & Galinsky, 2012). The influence on behaviour of the semiotic features of clothing was named 'Enclothed Cognition' and has received some support in other areas. For example, when participants were told they were wearing a nurse's scrubs compared with a cleaner's tunic, they reported higher feelings of empathy after reading a fictional account of a person who had just experienced a break-up (López-Pérez, Ambrona, Wilson, & Khalil, 2016). Further, clothing has been shown to affect self-perception. In a study on the effect of attire on self-rating, participants wearing either formal or casual outfits used an adjective list to describe themselves. Those in the formal condition rated formal words as more descriptive of themselves than those in the casual condition.

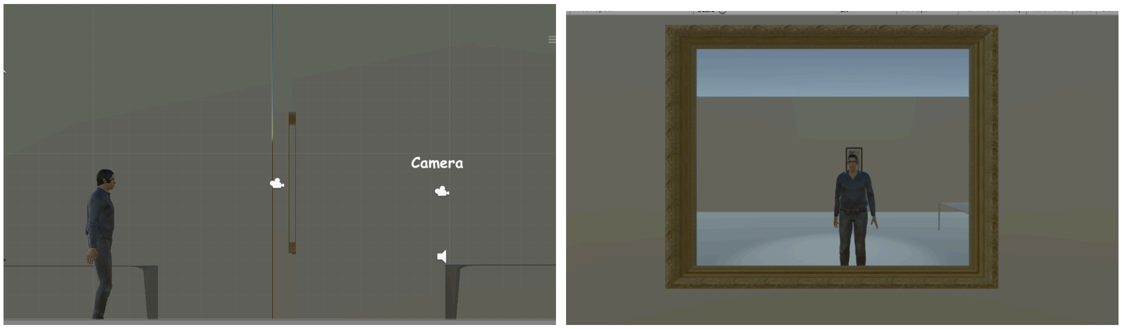
[19]This effect was not found in a larger group of student football enthusiasts however, and in order to find significant effects, the researchers had to look at individual play trials which was not performed on the group of referees, and was presumably not part of the original analysis plan.

In the laboratory studies mentioned above reality was somehow altered, whether through false heart rate feedback, or presenting participants with the illusion of choice. The possibility that attitudes or behaviour could be manipulated by *breaking reality* using virtual environments was the underlying assumption of the series of studies upon which the Proteus Effect is based (Yee, 2014, pp. pp142–143).

Proteus Effect studies

The previous sections have demonstrated how psychologists have manipulated various behaviours using physical stimuli such as clothing. In the Proteus Effect, attempts were made at manipulating behaviour using virtual stimuli, by altering self-perception using a body swapping illusion. In a typical experiment, participants wear a head mounted display and observe an avatar through a virtual mirror. The participant then ostensibly completes a task as the avatar, and performance was compared across conditions. In this thesis, the | notation will be co-opted from probability theory such that $Stereotype|Avatar$ indicates the evocation of a stereotype given an avatar. For example $Aggression|Height$, means the evocation of the stereotypical attribute $Aggression$ given the $Height$ of the avatar. In the first study from Yee et al. (2009), people embodied either an attractive or unattractive avatar. While still in the virtual environment, participants then engaged in a conversation with an avatar (controlled by a confederate) who was of the opposite sex. Participants in the attractive avatar condition maintained a smaller interpersonal difference with the confederate and disclosed more personal information during the conversation than those who embodied the unattractive avatar, which was interpreted as a $Sociability|Attractiveness$ effect.

The Proteus Effect has also been shown to influence behavioural economics tasks. In the ultimatum game, one player is given a commodity and told that they have to share it with a second player. If player 2 accepts the offer then



(a) Side view. The user looks in a ’mirror’ where an avatar tracks their movements. Yee used a ’fake room’ in 2007, however technology is now vastly superior and the same effect can be achieved by simulating light reflections.

(b) Front view: The participant looks in a mirror and gets to know their ’new body’

Figure 2.4: Mock up examples of Yee’s experiments made in Unity by The Author

both participants receive their cuts, whereas if the offer is refused neither player receives anything. When participants were immersed in a virtual environment, people who appeared taller accepted fewer and made more unfair offers compared to those who appeared shorter ($Aggression|Height$). In a subsequent study, the behaviours demonstrated in the virtual world were carried over into a repeated ultimatum game with a real person (Yee, 2007, Study 3). There is also some support for Proteus Effect from observational studies. Using BattleNet, which is the database for the MMORPG World of Warcraft, Yee, Ducheneaut, Yao, & Nelson (2011) found that virtual (rather than identified) gender correlated with healing behaviours in game which is constructed as a feminine role in such contexts, and has the activation $Healing|Gender$.

Priming or Proteus

A counter argument to the mechanism behind the Proteus Effect was proposed by Peña, Hancock, & Merola (2009) who suggested that participants were simply being primed by their avatars rather than adopting behaviours associated with the relevant stereotype. They based this assertion on a series of studies that demonstrated that attitudes and cognitions could be modified as well as behaviours; a finding that the authors suggested fitted priming theories more consistently. In their series of studies, participants used avatars in black vs white uniforms (Study 1), or in Ku Klux Klan robes vs a doctor's uniform (Study 2). Participants in Study 1 whose avatars were wearing black outfits reported greater intentions to attack unarmed characters, as well as other aggressive measures, and less group cohesion, suggesting ($Aggression|Clothes$). As with the Frank & Gilovich (1988) studies, black was assumed to carry aggressive symbolism, and Peña et al suggest that this symbolism affected their participant's attitudes. In Study 2, Peña et al. (2009) reported that Thematic Apperception Test[20] stories were more aggressive, and lower in *need for affiliation* in the participants who had KKK avatars than the transparent control avatars. The doctor's avatars did not differ from either condition. The authors criticise the self-perception element of the Proteus Effect based on the fact that it could not explain the inhibition of positive thoughts reported in their study. They stress that SPT can explain the aggressive attitudes from negative appearance, but a spreading activation explanation from the automaticity priming hypothesis (Bargh, 1994) offers a better fit.

In response to these findings, Yee & Bailenson (2009) reported a study in which embodiment was manipulated as one of the independent variables. In a partial replication of Study 1 from Yee et al. (2009), participants either embodied an avatar (mirror condition, i.e. had control over it) or saw a playback of the previous participants actions. They reported that participants in the mirror condition who had attractive avatars chose more attractive partners in a mock dating task then those in the unattractive-playback condition. They also found that participants in the mirror condition with unattractive avatars overstated their height in a mock dating website task compared with those with attractive avatars in the attractive condition, suggesting a need to compensate for their facial appearance. However, it is worth noting that this study was also an example of an existing effect not replicating. There was no difference between the attractive

[20]Originally designed as a measure of *Unconscious Fantasies*, the TAT involves the test taker using a series of images to tell a story, which is then interpreted. In its later uses, it was used as a measure of personality, values, and motivations (Murray, 1943; Vane, 1981).

and unattractive avatar conditions in terms of interpersonal distance maintained with a confederate avatar[21]. Yee & Bailenson (2009) suggest that the difference between the playback and mirror conditions in their study support the theory that embodiment is more important than observation in the Proteus Effect, and that self-perception is a more appropriate explanatory model than automaticity priming since the latter would predict no difference between conditions.

Further Demonstrations of the Proteus Effect

There have been surprisingly few attempts to investigate the effects of avatar embodiment on health behaviour. As will be discussed in greater detail in Chapter 4, the Proteus Effect has been applied mostly to exercise behaviours by varying avatar body shape (Joo & Kim, 2017; Li et al., 2014; Peña, Khan, & Alexopoulos, 2016; Peña & Kim, 2014), but at least two studies investigated the effect of body-shape on diet-related attitudes (Kuo, Lee, & Chiou, 2016; Verhulst, Normand, Lombart, Sugimoto, & Moreau, 2018). This may be because body shape holds more stereotypical information about performance and health than other attributes, however focussing on appearance in interventions may also have negative effects on user's esteem.

The following paragraphs will discuss the other, non-health related applications of the Proteus Effect, to demonstrate how the theory has been developed, and provide support and critique of the effect as a whole.

Under the Proteus Effect theory, people make inferences about their digital selves and act according to the traits that they believe others ascribe to the representation. This has been demonstrated in a wide range of scenarios and behaviours.

Fox, Bailenson, & Tricase (2013) investigated whether embodying objectified avatars could affect measures that correlated with self-objectification.

[21] It is unclear how this study is represented in the Ratan et al 2019 meta-analysis - The Author cannot determine whence the effect size of $r = 0.27$ is computed.

While immersed in a virtual environment, women observed an avatar that was self-similar or dissimilar and wearing objectifying or non-revealing clothing through a virtual mirror and had a conversation with the avatar of a male confederate. Women whose avatars were wearing sexualised clothing had reported more body related thoughts during a written task than those whose avatars were wearing less revealing outfits. Further, women who embodied self-similar avatars also rated higher on rape myth acceptance scales, particularly when the avatar was sexualised. Fox et al. (2013) compared their finding with previous research that found women who tried on swimsuits listed more body related thoughts compared with those who tried on a sweater (Quinn, Kallen, & Cathey, 2006).

Studies have also demonstrated that embodied avatar attributes can affect perceptions of physical objects. Obana, Hasegawa, & Sakuta (2017) found that participants evaluated the weight of an object as heavier when the tracked arm that they saw in a virtual environment was muscular compared with participants who had a less muscular arm ($Strength|Muscularity$).

The Proteus Effect has also been applied to performance-based tasks and implicit bias. Participants embodying an avatar that looked like Albert Einstein performed better on a test of executive functioning than those who embodied an avatar closer to their age (Banakou, Kishore, & Slater, 2018). Further, those in the Einstein condition showed less implicit bias against older individuals than those in the younger condition. Other studies have included creative ideation tasks as a dependant variable in which participants were body scanned wearing one of three outfits: office-wear, the participant's own clothes, or an artist's smock (Rooij, Land, & Erp, 2017). The results were fairly confusing. Creativity was rated highest in the control condition, followed by the creative avatar and then the work outfit, suggesting that alternative digital outfits reduced ratings on their measure of creativity. Although the authors labelled this effect the creative proteus effect, the only aspect that varied within the avatar was clothing that the participants had worn previously. The avatar was essentially a rigged 3D image of the participant wearing different outfits. That is, 'digitally mediated enclothed cognition' is possibly a more accurate description of the study.

Effect Size Estimate

In a recent systematic review and meta-analysis, the overall effect size from Proteus Effect studies was estimated (Ratan, Beyea, Li, & Graciano, 2019). In a sample of 46 studies, the authors report an overall small to medium effect size of $r(44) = 0.24$, and sub-group effect sizes of ($r(25) = 0.23$) for the effect of avatar appearance on behavioural, and ($r(17) = 0.26$) on attitudinal outcomes. Although the authors do not claim that this was part of their selection criteria, they seem to have been motivated by the p-value of the outcome. They state that in cases in which there were multiple dependent variables, only one was used. In Pilot Study 2 and Dissertation Study 1 from Yee's dissertation, there were three trials in the ultimatum game task[22]. In Pilot Study 2, the first trial yielded significant, whereas in the replication (Dissertation Study 1) the second trial was the significant outcome. In the meta-analysis, rather than choosing one outcome and sticking to it, only the significant effect sizes were selected. Arguably the Virtual Self Modelling studies conducted by Fox and Bailenson should not have been included since they do not involve avatars, rather animated models were used. Further, Fox's thesis has a number of failed replications throughout (Fox, 2010) which were not were included in the meta-analysis; and the interaction reported in Fox et al. (2009a) ought not have been included since the type of virtual representation was not included in the analysis (the study only reported a participant gender × presence interaction). Finally, some of the studies reported were analysed using Bayesian methods (Banakou et al., 2018), and it is not clear how the authors accounted for this in the analysis.

[22]Participants negotiate a *pot of money* with a confederate. If an offer is accepted, each person receives their cut of the pot, if it is declined neither party receives anything.

Criticisms

There are some methodological flaws with the original Proteus Effect (and related theories) studies that will be outlined below.

Auxiliary Theories In a pre-registered replication of the Adam and Jalinsky (2012) study, doubt was cast on the enclothed cognition effect (Burns, Fox, Greenstein, & Montgomery, 2019). The latter study had three times the participants, and trials, as the original and did not find a significant effect of lab-coat on Stroop task accuracy, or reaction time. The data were tested using equivalence testing in which the hypothesis that the data falls outside a region of equivalence between zero and the smallest effective of interest (SESOI) is tested using Two One-Sided Tests (Daniël Lakens et al., 2018). The authors report that the effect of lab coats on accuracy did not differ from zero, and that the 90% confidence interval of the observed effect fell outside of the smallest effect size of interest (Burns et al., 2019). The authors conclude that they believe clothing may change behaviours, and attention towards certain stimuli, but feel that their study brings into question Adam and Galinski's claim that clothing may affect selective attention. They cite the work mentioned above by López-Pérez et al. (2016) as demonstrative that the symbolism of clothing may affect helping behaviour, but only when meaning is attached to it.

Sample and Effect Sizes The studies had very small samples. In Yee (2007), Study 1, 32 participants were sampled for an independent samples t-test. In order to achieve 95% power (i.e. correctly reject the null hypothesis 95% of the time, given that there is a population effect, and the study were to be repeated indefinitely) a Cohen's d of 1.2 would be required. This is an unrealistically large effect size for a social psychology study, in which there are a great many mitigating variables involved (Richard, Bond Jr, & Stokes-Zoota, 2003). Indeed, the average effect size reported by Richard et al. (2003) based on over 100 years of social psychology was $r = 0.21$ with a standard deviation of $r = 0.15$, which, when converted into standardised mean difference is $d = 0.43$[23] . In Yee's results, an effect size of $d = 0.4$ is reported for interpersonal distance and $d = 0.38$ reported for self-disclosure. However, upon calculating the effect size using the reported t-value using the compute.es package in the R programming language (R Core Team, 2017; Re, 2013) , $d = 0.86$, $95\%CI = [0.1, 1.61]$ and $d = 0.78$, $95\%CI = [0.04, 1.54]$ respectively were derived. Suspicion about the accuracy of the estimate can and should be aroused when small N studies have such large effect sizes and wide confidence intervals, since random variation in individual participants is given too much weight within the study and would be expected to even out under larger sample sizes (Button et al., 2013). The size of the confidence intervals around the effect size suggests that sampling was inadequate, since these shrink as a function of a reduction in the standard error of the mean, which itself reduces with more observations. Further, it is worth noting that neither of these effects replicated in a (more complex) subsequent replication (Yee, 2007, p. pp78; Yee & Bailenson, 2009). Taken together, the large smallest required effect, the improbable calculated effect sizes, wide confidence intervals, and failures to replicate suggest an instance of the "Winners Curse", followed, ironically, by the "Proteus Phenomenon" (Button et al., 2013). These related phenomena occur when the desired p-value required for publication (typically <0.05) is driven by an artificially large effect size. Although this may result in papers in journals with high impact factors, it also means that direct replications (which are becoming increasing popular in psychology) are more likely to fail, and sample sizes are more unlikely to be sceptically planned, since studies are expensive and why overpower one study when you could use the resources for several (Button et al., 2013). This means that in the long run, a set of effect sizes will regress to the (potentially significantly) smaller unknown population effect size. This shrinking of meta-analytic effect sizes is called the Proteus Phenomenon for the same reason

[23] Note that this 100 years includes many decades of 'publish or perish' incentive structures, and publication bias is likely to have inflated these estimates considerably(Chambers, 2017; Ioannidis, 2005).

that Yee named his theory the Proteus Effect (i.e. after the shape shifting Greek God).

Figure 2.5 presents a scattergram of the effect sizes from Ratan et al. (2019) as a function of sample size. There is a negative correlation of r = -0.52 [24]. The authors of the meta-analysis discuss this in terms of publication bias and provide a fail-safe N as an adjustment. However, if a true effect existed (the authors suggest $r = 0.23$), larger samples ought to capture a larger effect. In this case, it appears that effect size is acting as the Greek god by changing its form for each who observe it!

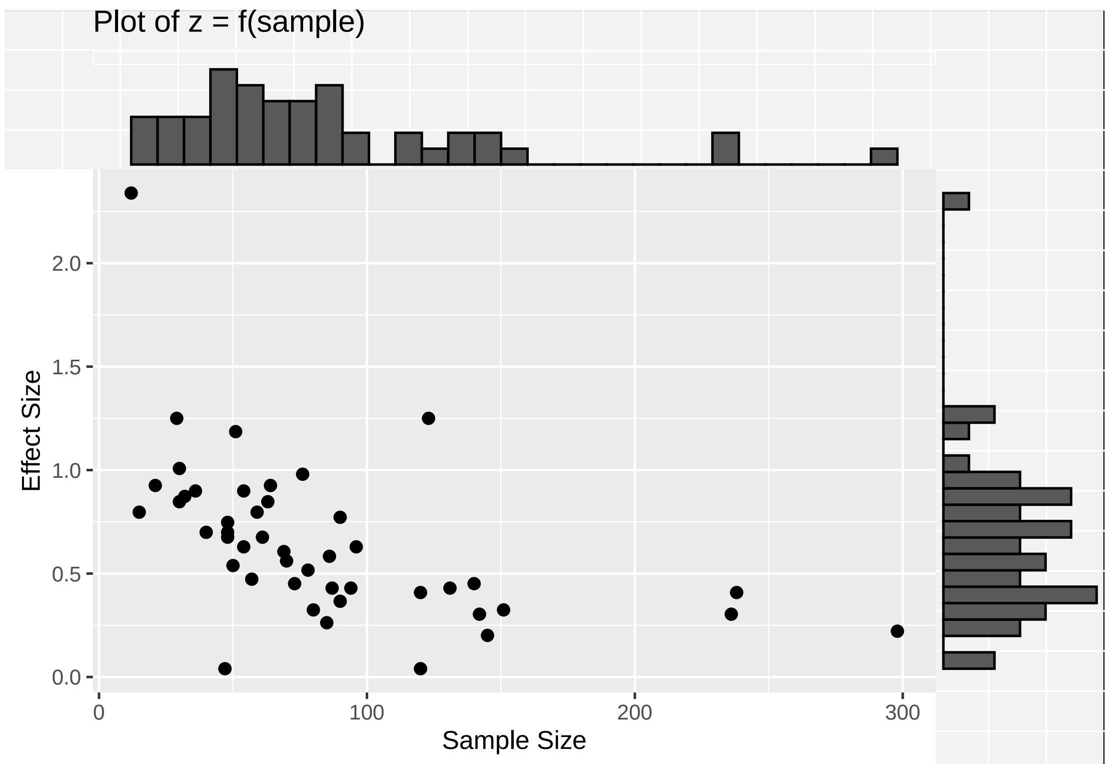


Figure 2.5: Scattergram and marginal histograms demonstrating the negative correlation between sample size and effect size in the Ratan et al (2019) data set.

Lack of Avatars in two studies The next issue with the original Proteus Effect studies concerns Experiment 2 from Yee et al. (2009). In Study 1 participants observed themselves in a mirror and were given the opportunity

[24]FWIW: $t(44)$ = -3.99, $p = < .001$. It says ”No-P-Values”, I’m allowed to have one.

to draw inferences about themselves from the physical features of the avatar. In Studies 2 and 3 there was no virtual mirror and participants were just made taller or shorter as relative to the game world. This means that the only cue available to them was the confederate avatar, suggesting that the inferences that informed their behaviour were drawn from the confederate rather than the self. There is some observational support for the shorter person in a dyadic competitive situation (in this case navigating a narrow path or corridor) to yield to the taller person (Stulp, Buunk, Verhulst, & Pollet, 2015), and height has been shown to affect interpersonal dominance in virtual reality however this study had a very small sample (Pazhoohi et al., 2018). As such, it is unclear whether the Proteus Effect is the best explanation for what was observed in Studies 2 and 3. Further, the hypothesised effect - that there would be a significant difference in the first split of the ultimatum game - was not observed. It was the second split that was "highly significant", which was not in line with previous research but was explained *ad hoc* [25], based upon discussions with the research assistants who stated that participants seemed to be "testing the water" . This is a fair assessment but does not constitute a confirmatory hypothesis and ought to have been replicated before being reported as such. In fact, an attempt at a replication of the experiment was run in a later study (Yee, 2007, Pilot Study 4, pp46-51), however in this subsequent study the significant split was the first, in line with the original hypothesis based on prior research that suggests splits should become less aggressive over time (Bolton, 1991) . As such the reported effect has considerable noise which may again be due to small sample sizes and the reliance of large effect sizes to carry the significant results (both required an effect of $f > 0.5$ to reach 95% power, where $f = 0.4$ is a 'large' effect size; Cohen, 1992).

Essentially, although there is *money in the bank*[26] for the Proteus Effect as a theory, a degree of scepticism about the effect is warranted. In Chapter 3, the benefits of Bayesian statistics for including scepticism in statistical models will be

[25] ad hoc_3 by Lakatos' classification, see page 67.

[26] Term used by Meehl (1990) in reference to Lakatosian defence to describe the accumulation of evidence in support of a theory.

discussed as a method for handling this uncertainty.

Stereotypes and Stigma

At its core, the Proteus Effect relies on stereotypes for inferences about the self to be drawn. Stereotypes are pervasive, often inaccurate beliefs about a group of people (Jussim, Crawford, & Rubinstein, 2015; Operario & Fiske, 2004). They generally emerge when there is a dearth of individuating information about a person (Carlston, 1994; Kunda & Thagard, 1996). Fiske & Taylor (1991) conceptualise stereotypes as existing in individual minds and consisting of associations between a group and an attribute, such as that between *gamers* and *laziness* (Kowert et al., 2012). Once learned, these associations become obdurate, and are activated automatically on exposure to a group or group member. Falbén et al. (2019) note that stereotypes facilitate expectancy-consistent processing, i.e. processing of information and subsequent decision-making based upon consistent information is faster than when information is inconsistent with expectancies and confirmed this across occupational and trait levels of information, including a variety of tasks. Stereotypes have been reported across a variety of broad groups, including gender (Diekman & Eagly, 2000) and race (Katz & Braly, 1933); as well as specific groups such as 'female politicians' (Schneider & Bos, 2014), student athletes (Wininger & White, 2008), and online gamers (Kowert et al., 2012).

A common approach to identifying stereotypical traits in different groups is to ask groups of individuals to list words that they feel stereotypically represent the group in question. This method was pioneered by Katz & Braly (1933), who asked a small group of Princeton students (N = 25) to list words they associated with various ethnic groups (Jewish, Black, Irish, and Chinese people). They then asked a second group of 100 Princeton students to select the five most relevant words that described the groups. They report the 12 most selected words from each group. The checklist method has been repeatedly used and validated against other methods with similar findings identified between cultures (Schneider & Bos, 2014; Stephan et al., 1993).

Other researchers have adapted the classic Katz and Braley method to use polychotomous rather than dichotomous responses. Kowert et al. (2012) used a 7-point scale on the applicability of the words related to online gamers. This allows more variation in the ratings of words and means that statistical methods such as factor analysis may be used to identify latent stereotypical constructs that generate associations with the words (Cuddy et al., 2009; Kowert et al., 2012). Kowert et al. (2012) identified stereotypes of online gamers based on factor loadings of numerous words. These included *idleness, attractiveness, sociality, and popularity.* In this thesis, a combination of methods was used in identifying stereotypes of athletic people. Initial word lists were generated in a similar manner to Katz and Braley, but the method of stereotype verification was similar to that of Kowert et al. (2012). This allowed a more sophisticated analysis of the latent structure of the stereotypes of interest. Full details of how this was achieved in the current study are provided in Chapter 6.

Information Structure of Representations

For the purposes of this PhD, Associated Systems Theory was employed to model the structure of stereotype representations (Carlston, 1994). In AST, representations of people are built from four interacting systems: Visual; Verbal; Action; and Affective (Schleicher & McConnell, 2005). These vary in abstract-concrete and internal-external dimensions. Abstract systems (affective, verbal) are deemed to be stable, whereas the concrete systems such as the way a person acts and looks will change with time. A representation of this system is presented in figure 2.6.

Combinations of each of these systems result in the encoding of different information. AST predicts that stereotypical evaluations of a person will be activated when there is a paucity of information available, such as visual

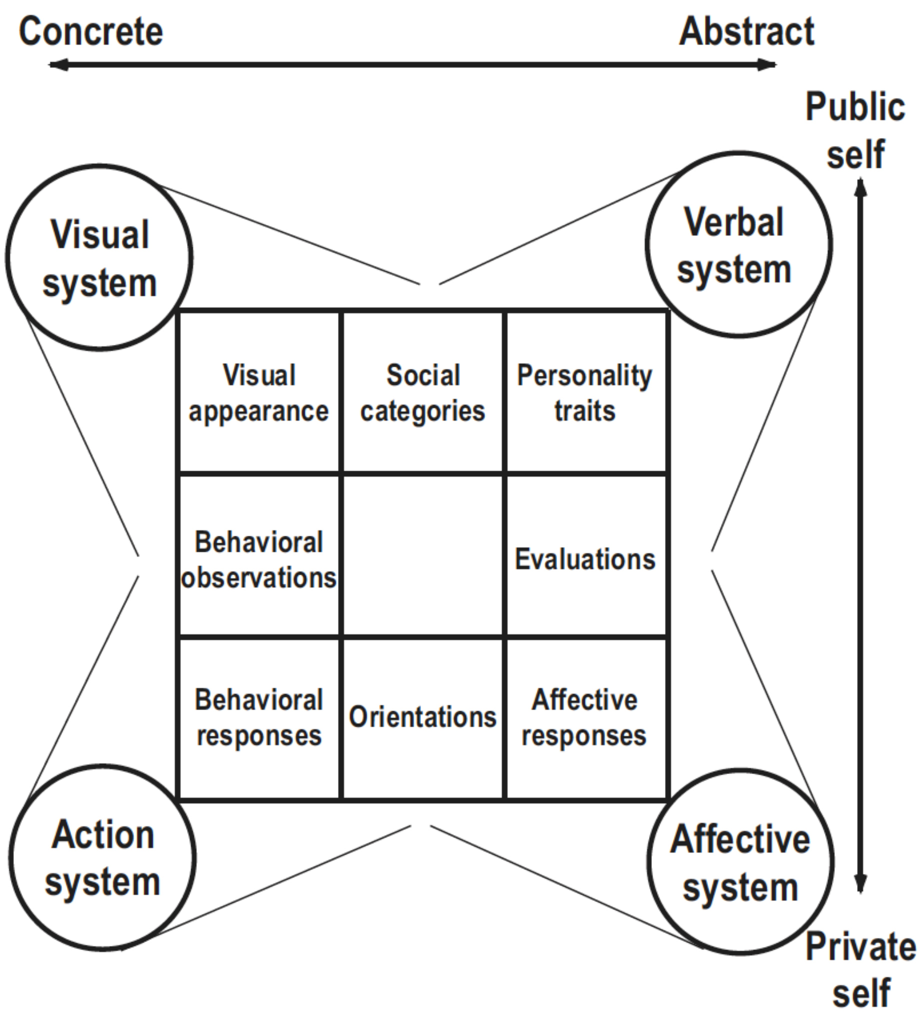


Figure 2.6: Representation of data about individuals as modelled by Associated Systems Theory, Schleicher & McConnell (2005).

information with no behavioural, or personality data. The more information available under this model, the more likely an accurate representation is to be created, and correct inferences to be drawn.

Taking this perspective, the Proteus Effect can be described in terms of concrete information (visual appearance of the avatar) being used to infer abstract information (traits) in the AST framework.

This is relevant to the current thesis, since when an avatar is first observed, people draw inferences from the information available to construct a representation. If only limited (in this case visual) information is available, then the representation will be formed based on stereotypes (e.g. the avatar is overweight, so it must be lazy). Likewise, if only behavioural information is available (e.g. "the person can run 100 meters in 10 seconds"), then inferences will be made about the absent systems, e.g. "this person must be athletic, dedicated, fit". Indeed in a later study, participants will be presented with avatars with different amounts of information regarding behaviour with the expectation that provision of fitness-related behaviour will reduce the response to words associated with idleness (or low responses to words associated with fitness) when the participant is presented with a larger-bodied avatar (Chapter 9).

Stereotypes and the Proteus Effect

Perhaps the most interesting Proteus Effect study is also the least well-known. In the last study of Yee's PhD thesis, attempts were made to investigate the Proteus Effect when no stereotypes were associated with an avatar (Yee, 2007). This approach involved using the Katz and Braly (1933) list generation method to identify stereotypes associated with people who were improbably tall (8.6 meters), and improbably small (36cm). In the absence of stereotypical information based on these hypothetical groups, Yee predicted that participants would fall back on general stereotypes associated with height. However, rather than the participants displaying extreme behaviour to go with their extreme appearance, Yee predicted that participants would compensate by behaving incongruently with these behaviours, by increasing their offers (i.e. being less aggressive) when they were unfeasibly large, and decreasing the size of their offers (being more aggressive) when unfeasibly small. The results from the study did not support the hypothesis, however the robustness of the approach was a strength of the study. In the current thesis, some of the methods used to identify stereotypes in groups and apply these to behaviours were adopted.

Body Stigma, Body Image, and Health

A widespread example of negative stereotypes applied to appearance is *weight stigma*. This section will discuss research from *Body Image Psychology*, a subfield of Psychology that deals with the relation between a person and their body (Grogan, 2017). This field transcends distorted bodily perceptions, and researchers in the field broadly investigate how personal, social, and environmental factors affect the perception of and satisfaction with the body. Since avatars temporarily represent the player's physical form in a virtual environment, there are a number of theories from the body image literature that are relevant. For instance, some researchers have discussed the avatar design choices in terms of a compensatory function in individuals with lower body satisfaction (Cacioli & Mussap, 2014). Before delving into this side of the avatar research, some discussion of what body satisfaction is, and how it varies between individuals and groups is warranted.

Body image is a person's perceptions, feelings, and evaluations of their external appearance (Grogan, 2017). Because of pressures to conform to normative appearances in developed countries (e.g. slender for women; muscular for men), people may compare themselves to an idealised norm, resulting in a discrepancy between their actual body and an internalised ideal. This discrepancy can be used as a measure of body dissatisfaction, which is related to a variety of negative outcomes such as eating disorders, controlled and unhealthy eating

(Neumark-Sztainer, Paxton, Hannan, Haines, & Story, 2006) and depressive illnesses (Paxton, Neumark-Sztainer, Hannan, & Eisenberg, 2006). Although there are no body image measures in this study *per se*, the research on responses to similar stimuli (e.g. images of muscular or idealised figures) is pertinent for predicting how stimuli will be responded to as a function of how a person feels about their body, which may be of interest in future studies.

The existence of negative stereotypes and stigmatisation towards plus-sized individuals has been well researched since as early as the 1980s. Robinson, Bacon, & O'reilly (1993) first started exploring the types of associations that the general US public make with people with larger bodies in 1984 (Robinson et al., 1993, p. p470). Attendees at a driving license bureau were asked to list adjectives to describe people who are fat. These words informed the development of a weight stigma measure, or 'Fat Phobia scale' in 1993. Robinson et al. (1993) report 6 factors (50 items) to weight stigma: Undisciplined/Inactive/ Unappealing; Grouchy/Unfriendly; Poor Hygiene; Emotional/Psychological Problems; Passivity; Stupid/Uncreative. These 6 factors were reduced further in 2001 by Bacon, Scheltema, & Robinson (2001), who created a 14 item *semantic differentiation* scale, in which participants rated how close a person fell between two antonyms (e.g. lazy, industrious).

The presence of weight stigma is also well documented both in the general (presumably North American) population (Puhl & Brownell, 2006) and in fitness professionals (Ntoumanis, Guerrero, Gadeke, & Thøgersen-Ntoumani, 2018). It is also consistent across many post-colonial countries including Australia and Canada (R. M. Puhl et al., 2015), and Brewis, SturtzSreetharan, & Wutich (2018) present evidence that concern with body size and anti-fat attitudes are spreading to South America, Russia, India, and China, and that this may pose a health issue. Arguably, the stigma associated with larger bodies has led to a focus on appearance related interventions. Many of the avatar appearance-related interventions reviewed in Chapter 4 cite obesity (as in the *Obesity Crisis*) as the driving factor behind the research, and some earlier studies on the use of Virtual Reality as an intervention focused on obesity as a target (Fox & Bailenson, 2009). Elsewhere, a rise in obesity and heart disease in many Western cultures is cited as the justification for researchers to explore ways to encourage larger bodied people exercise (Lamboglia et al., 2013). Indeed, some exergames *reward* players with reduced body size for exercising (e.g. Wii Fit), or provide players with athletic-looking representations of themselves, leading to a reinforcement of the 'thin is good' mentality (e.g. Chapter 5), when there is evidence to suggest that a more physical activity interventions may be more effective when the message is 'functional is good' (Mulgrew, McCulloch, Farren, Prichard, & Lim, 2018).

The degree to which body size ought to be part of the discussion in health may be further criticised through an exploration of the relationship between fitness and disease (Monaghan, 2005). As mentioned previously, actual cardiovascular fitness appears to influence the risk of all-cause mortality. Fitness predicted all cause and cardio-vascular disease related mortality in a prospective study (Blair et al., 1989), and this was found in both larger and smaller bodied individuals (Lee, Blair, & Jackson, 1999). What is pertinent is that a large scale study on health risks in adults with 'healthy' body mass indices, compared with those with *unhealthy* BMI, found that sedentary behaviour was more of a risk factor (i.e. a larger differential of risk increase) in the healthy BMI condition (Mainous III, Tanner, Rahmanian, Jo, & Carek, 2019). This has been further supported in a meta-analysis finding that mortality was comparable between larger and smaller body types with similar levels of fitness; this finding is called *The Obesity Paradox* (Barry et al., 2014).

Although body size is influenced by sedentary behaviours, there is a strong link between low physical activity and sedentary behaviours and heart disease regardless of body size. The stereotypical causal chain is that people have larger bodies, ostensibly because they are lazy and undisciplined. Indeed, some researchers argue that heart disease is caused by larger body sizes, rather than unhealthy behaviour (e.g. Lyons & Hatkevich, 2013). This message leads to the stigmatisation of larger bodied people, and perhaps the perceived vindication of those who are sedentary but not overweight, leading to the false belief that they are 'safe', and exercise is not needed. A stronger case is that obesity may be related to sedentary behaviours, but frequent movement is required to maintain adequate fitness regardless of a person's body size, and prolonged sitting should be avoided.

An element of the experience of weight stigmatisation includes the erroneous belief that one is *too fat to exercise* (Ball, Crawford, & Owen, 2000). Internalisation of anti-fat social attitudes has been associated with exercise avoidance (Vartanian & Novak, 2011), and experience of weight stigma has been associated with body dissatisfaction, and reduced desire to exercise regardless of body type (Vartanian & Shaprow, 2008). This is a problematic barrier, since as previously mentioned, exercise and fitness are protective regardless of body size. One of the aims of this PhD is to investigate whether weight stigma may be reduced using avatars be digitally representing larger bodied fit individuals.

Is there an athletic stigma?

If larger bodies are socially undesirable, then are athletic bodies desirable by extension? There is very little research on stereotypes related to fitness and athleticism. Despite this, there is an assumption that people who exercise regularly are motivational for *normal* to *plus-sized* individuals is common. For example, the *Encyclopaedia of Body Image and Human Appearance* has a brief entry on exercise and body image, making the general assumption that people who exercise are held in high regard :

> Exercise may be the most revered of all the body-changing strategies. Indeed, in contemporary society, exercise is a symbol of determination, hard work, and control. Exercisers are stereotyped favourably and

> rarely criticised for exercising too much or being too committed to their exercise regime. One might argue that exercise dependence is an acceptable addiction in modern society. (Ginis & Bassett, 2012)

The presence of societal (*out there*) stereotypes towards fit and athletic individuals is reflected by media depictions of the ideal body (Grogan, 2017), however, it has received little attention in the stereotype literature, and it is unclear upon what Ginis and Bassett's claim about favourable stereotypes was based. Some early studies on student athletes identified a 'dumb jock[27]' stereotype (Adler & Adler, 1987; Sailes, 1993), but this is far from favourable. Further work was conducted on intersectional stereotypes of black athletes who were deemed to be 'lazy' but 'naturally gifted' in their abilities, with intelligence being a secondary attribute (Czopp, 2010; Moskowitz & Carter, 2018). These negative stereotypes have been identified in student populations in which participants were asked to evaluate the academic prowess of student athletes. For example, in one study, students reported lower academic expectations for student-athletes than those who were not student-athletes (Wininger & White, 2015). The stereotype associated with black-athleticism is 'Naturally athletic' (Sailes, 1993), in which Black athletes are regarded as having not tried as hard, with their skill being attributed to their natural ability. This *racialised performance* interpretation of black athletes features prominently in earlier discussions of racial stereotypes (King, 2004), and is associated with infantilisation, and dehumanisation by white people of athletic *others*, including black people and native Australians (Coram, 2007).

Both of these stereotypes (natural ability and low academic prowess) have received some behavioural evidence in the stereotype threat literature[28]. For instance, black participants who received a 'natural golf ability' prime had a lower performance than those who did not (Stone, Lynch, Sjomeling, & Darley, 1999); and discussions with a large group of college athletes, Simons, Bosworth,

[27] According to the Merriam-Webster dictionary, a jock is "An Athlete, especially a school or college athlete"

[28] Although the general body of Stereotype Threat research is up in the air at the moment due to failed replications.

Fujita, & Jensen (2007) found that a low percentage received positive comments from faculty, and public stigmatisation of the group 'athletes' was very common. They also report that athletic stigma exists across sexes and races, although black students reported higher stigma, and the 'double stigma' of being black and an athlete. It appears as though there is a degree of stigma associated with athleticism, although perhaps not as diffuse and severe as weight stigma, but this has mostly been investigated with reference to student athletes, and mostly in the United States.

Relevance to the Proteus Effect

The Proteus Effect holds that when embodying a digital representation (or avatar) that has attributes strongly associated with a stereotype, people will adopt behaviours congruent with that stereotype.

Wininger & White (2015) describe the distinction between 'ascribed' belief and 'identifying' features of stereotypes. An ascribed belief is one about an attribute assigned to members of group, e.g. 'unintelligent' may be ascribed to the group 'jocks'; identifying features refer to the attributes that would lead to an individual being categorised as a member of the group. For instance, physique, behaviour, and attire are identifying features. To the best of The Authors' knowledge, no work has been conducted on isolating identifying characteristics related to athletic individuals and exploring the ascribed beliefs associated with these alone[29]. The one of the objectives of this thesis is to induce evaluations based upon ascribed beliefs using identifying features.

Studies on the Proteus Effect have shown that in-game behaviours may be manipulated when playing a game with particular avatars (Peña et al., 2016; Peña & Kim, 2014; Yee et al., 2011), or during immersive virtual reality experiences (Fox et al., 2013; Yee et al., 2009). This is important because

[29] *Author Note*: When referring to myself as the author of the thesis, I will use the title case. I would like to apologise in advance if this comes across as pretentious - however I frequently refer to authors of other papers as *the author*, and this is used to reduce confusion.

a pre-requisite of the Proteus Effect is that the player experiences a shift in self-perception prior to adopting the behaviour associated with the avatar, rather than simple priming which Yee & Bailenson (2009) argue was ruled out based on their experiment comparing embodiment of an avatar during the enacting of a scene, with the mere observation of a scene. The effect of avatar body size on physical activity has also been investigated by Peña and colleagues, who demonstrated that female waist (Peña & Kim, 2014), and male wrist (Peña et al., 2016) movements were lower when playing a Wii tennis game with an "obese avatar"[30], compared with those playing with a "normal-weight" avatar. Other studies have also demonstrated that both intentions to exercise and observed behaviour can be diminished when overweight avatars are used, leading some authors to call for only thin and muscular avatars to be available in exergame interventions (Joo & Kim, 2017; Li et al., 2014). These studies are discussed, *ad nauseum* in Chapter 4.

Behavioural Responses

The work of Peña et al (2014; 2016) suggests that exposure to athletic body types has a positive effect on health-related behaviours. However, there is evidence that this interaction may be more complicated. For example, Wasilenko, Kulik, & Wanic (2007) demonstrated that exposure to fit peers reduced exercise duration and body satisfaction in a naturalistic study. Additionally, unfavourable comparison of the self to the bodies of others was reported as a reason for gym non-adherence by women in an interview study by Pridgeon & Grogan (2012), suggesting that social comparison with an avatar is a potentially important factor for understanding the experience of using avatars with different body sizes. However, the relationship between exposure to fit-looking bodies, upward physical comparisons, exercise performance and body image related affect is possibly more complex. Halliwell, Dittmar, & Orsborn (2007) found that

[30] Their phrasing.

exposure to media images of muscularity increased negative body affect in men who did not exercise but had the reverse trend in men who did exercise. Pila, Barlow, Wrosch, & Sabiston (2016) demonstrated that men with more negative body evaluations participated in more exercise when they made body related upward social comparisons whereas the reverse was observed in women. Elsewhere, exposure to images of average rather than thin (Diedrichs & Lee, 2011) or muscular (Diedrichs & Lee, 2010) models was found to have a positive effect on body satisfaction in those who internalised social norms surrounding beauty. Presumably, inferences about these 'athletic others' are being drawn from somewhere in order to engage with in these comparisons, but from where is unclear.

As will be discussed below, behavioural responses to stereotypes are dependent upon two factors, susceptibility and diagnosticity, (Hilton & Fein, 1989; Steele & Aronson, 1995).

Stereotype Susceptibility

The elicitation of in-group related stereotypes in laboratory studies has been shown to affect performance in both positive and negative directions. Reaction to stereotype priming is named 'stereotype susceptibility' and may be seen as having three distinct categories: threat, boost, and lift. Stereotype threat occurs when a stereotype infers that an in-group is inferior in a task compared with an out-group; for example, priming participants with a stereotype about inferiority compared with white participants was associated with lower performance by African American participants (Steele & Aronson, 1995). Stereotype lift occurs when the stereotype threat prime is reversed, and an in-group is primed with the inferiority of an out-group (Walton & Cohen, 2003). Stereotype boost occurs when participants are primed with superiority of a task, for instance Asian American women primed with the stereotype of Asian Americans being good at mathematics performed better than Asian American women who were primed with the women being poor at mathematics stereotype (Gibson, Losee, & Vitiello, 2014; Shih, Wout, & Hambarchyan, 2015). Although not originally framed in these terms, the Proteus Effect could be discussed in terms of stereotype threat, boost or lift. This would require that:

1. Participants were aware that a stereotype for their avatar existed (C. E. Gibson, Losee, & Vitiello; 2014).
2. The use of an avatar that evokes these stereotypes.
3. The opportunity to perform a stereotype-relevant behaviour is available.

The internal representation may be a held stereotype (*in here* type), or a belief about the stereotypes held by others (*out there* type). This is distinct from the notion of *meta-stereotypes* which are in-group member's beliefs about the stereotypes held about the in-group by an out-group (Vorauer, Hunter, Main, & Roy, 2000), however it could be argued that the Proteus response is the result of a meta-stereotypical activation, since the mechanism is a function of how observers believe the subject should act.

Some unconfirmed potential corollaries to this list include the adoption of an avatar-as-self relation (cf. page 22), and that the experience induce embodiment and/or self-presence (cf. page 24). First, as discussed previously, Jaime Banks' taxonomy of Player-Avatar Relations varies on the distance from the self that the user places the avatar. It is unlikely that a person who sees their avatar as merely an object would adopt behaviours associated with that avatar; moreover, someone who sees the avatar as a symbiote or an other may act in a corresponding way, but for different reasons, such as conscious role playing (the Proteus Effect is assumed to be non-conscious). In which order the relations and behavioural adoptions stack would be important here. If the Proteus Effect is a truly automatic process that occurs early on in the avatar perception system, then relation-style would have no effect on behavioural modification; whereas if relation-style occurs before the

Proteus Effect, or is indeed a trait that is diffuse and has generalised effects across all avatar interactions, then this may nullify Proteus Effect-related behavioural interference. Second, presence is not discussed by Yee (2007), and is seldom included as a moderator. It might be that presence was taken for granted in the early VR studies, since the total immersion in a 3D environment can be quite persuasive (the brain *buys into* it). In studies where presence has been discussed, such as Fox et al. (2013), it was not found to be a significant moderator - but if virtual immersion is so effective, there may have been a restriction of the range of the presence measurement, leading to a reduced correlation between presence and the outcome . It is notable that presence was not discussed as a moderator in the meta-analysis by Ratan et al. (2019), and so it's importance as in the Proteus Effect is as yet unknown.

In the original set of studies (Yee, 2007; Yee et al., 2009), participants embodied unattractive or attractive avatars or experienced being taller or shorter than a confederate's avatar in a virtual environment. In the case of the second (height) study, the *aggressive* negotiating by the taller condition could be seen as a stereotype boost (increase by virtue of being taller and therefore more assertive in negotiations) or a stereotype lift (increase by the confederate being shorter and therefore less assertive in negotiations). In the shorter condition this could be seen in terms of stereotype threat. This would only be the case if implicit priming through sensory embodiment was successful, or if participants were explicitly primed beforehand (e.g. informing participants that shorter people are generally less apt with negotiation tasks) which did not happen in Study 2 or 4 from Yee (2007).

For studies in which the stereotype is unclear, such as Peña & Kim (2014) and Peña et al. (2016), the explanatory outcome is difficult to parse. Playing as an "obese" avatar could have a stereotype threat element to it or playing as a normative ("normal-weight", ostensibly athletic) avatar may have had a stereotype boost element to it. The only behaviours measured in these studies were the number of wrist and waist movements, a reduction in which is perhaps loosely associated with stereotypes of laziness in "obese" individuals. However, whether the number of movements was associated with motivation, energy, or desire to win remains unclear. A final study to consider is that of Li et al. (2014) which, to the best of the author's knowledge, was the only study to actually use a stereotype threat prime. In this study, no interaction was reported between the appearance of the avatar ("overweight" or "normal") and the use of a stereotype threat prime regarding higher BMI levels - although each condition affected the psychosocial outcomes independently.

The assumption in these studies was that athletic/normative and obese/overweight traits and behaviours exist on a single dimension with overweight people with a sliding scale on each trait that determines whether a person will be fit or fat.

A stronger test of the Proteus Effect, and/or stereotype susceptibility and body type would be to identify traits associated with stereotypes related to athleticism, and to having a larger body, and use these to infer appropriate behavioural responses. Using these predictions, participants would be given an opportunity to perform these behaviours whilst embodying avatars holding those physical attributes. Part of this test would involve exploring whether judgements of *fitness* and *fatness* are correlated, despite evidence that these are orthogonal constructs (Lee et al., 1999).

Diagnosticity

When casting inferences about individuals, stereotypes are generally employed when there is a drought of information available. That is, stereotypes are less likely to be employed when individuating information that is highly typical (i.e. relevant across many social judgements) is available (Hilton & Fein, 1989). Further, when individuating information is available, i.e. information that sets a person apart from a group, this is relied on in judgements (Jussim et al., 2015). The upshot of this is that more diagnostic information is associated with less reliance on stereotypes (Crawford, Jussim, Madon, Cain, & Stevens, 2011).

Virtual models provide a unique opportunity to explore this assertion. Since they are highly customisable, 3D models can be morphed into different shapes whilst keeping facial features constant and can be animated to perform any action. This means that the same model can be presented as athletic and muscular or plus-sized and anywhere in-between. Under the framework described above, increasing diagnostic information about individual 3D models ought to reduce the reliance of participants on stereotypes. For example, a plus-sized avatar could be animated to competently perform rigorous exercise. This behaviour contradicts the group stereotype for over-weight or obese individuals (*lazy, slow*) but provides diagnostic information about the plus-sized individual (*energetic, fast*), meaning that there ought to be less reliance on a stereotype for inferences on the individual and therefore fewer negative stereotypes ought to be elicited under this framework.

In summary, on the one hand athletes may be stigmatised as unintelligent; whereas on the other people who may have identifying features such as athletic physiques may be deemed motivational and the diametric opposite of the fat-related *lazy* stereotype. The latter interpretation was the one used in an Implicit Association Task study by Robertson & Vohora (2008). Historically, only one of these two classes (stigma and reverence) of stereotypes are investigated at a time. One of the purposes of the studies in this thesis was to explore the types of stereotypical inferences that are invoked when only information about physique is provided, since often this is the most salient information about fitness available on first impressions. Further, based on the aforementioned discrepancy in reported reactions to upward social comparison (e.g. Pila et al., 2016), there may be a distinction between a shared stereotype at the societal level propagated by agencies such as the media and governments (Haslam, Turner, Oakes, Reynolds, & Doosje, 2002), and cognitive-based individual level stereotypes which are thought to be used as heuristics within individuals (Fiske

& Taylor, 1991). This distinction is referred to as the *In Here* and *Out There* distinction (Kowert et al., 2012) .

2.6 Thinking Fast and Slow

A possible consequence of the provision of stereotypical information is that this irrelevant information is favoured over relevant facts. Later in the thesis (Chapter 9), a proxy for *hubris* regard was required to identify whether ambivalent stereotypes affected participants self-perception. Hubris was chosen because in Chapters 6-8 a stereotype associated with athleticism was arrogance. In line with the necessary requirements for the Proteus Effect outlined on page 52, this had to be achieved whilst the participant was being represented by the avatar. As a result, participants were asked to rate *their own* predicted performance on a second athletics track. This approach was based on the concept of base rate neglect (Kahneman & Tversky, 1973), the background to which will be discussed below.

System 1 and System 2

In the book *Thinking Fast and Slow*, Daniel Kahneman (2011) describes two systems for human cognition, which he names System 1 and System 2. System 1 produces fast, intuitive judgements and is optimised for economy of thought, and System 2 is engaged for slower, more methodological and rational processes. According to Kahneman (2011), System 1 is more prone to the making of erroneous judgements based on irrelevant information such as appearance, making the activation of stereotypes a product of System 1 (Kahneman, 2003). Kahneman (2011) describes the principle of *What You See is All There Is*, which is a feature of System 1 whereby inferences and predictions are made using only the information available at the time. It could be argued that the Proteus Effect (at least in the virtual reality implementations of the effect) is an example of this principle. When a person's sensory experience is manipulated such that they feel taller or shorter, or when they look in a virtual mirror they see an attractive or unattractive avatar, the Proteus Effect suggests that a person's self-perception is shifted, and their behaviour alters. This shift is not a rational act, but nor is it a delusion. Essentially, the effect relies on the *What You See Is All There Is* heuristic of System 1. Were the person to engage their System 2 processes, they would be able to identify the sensory experience of the virtual environment and body as synthetic.

System 1 uses a set of heuristics to expedite decision making (Tversky & Kahneman, 1974). These include: availability, or the presumed frequency of an event based on previous experience; anchoring, whereby estimates do not veer far from a given starting point; and representativeness, or the degree to which a target is similar to a stereotype of a group when making a judgement. This latter heuristic is the most relevant for the current project, since it relates to stereotypical judgements. Under the representativeness heuristic, a person's estimate of the probability that a target having an attribute is informed by irrelevant information. In this thesis, the irrelevant information is the physique of the avatar.

2.7 Summary

This chapter has highlighted the theoretical framework behind behavioural change using avatars, as well as some potential moderators of the effect. In brief, under the Proteus Effect:

1. Altering self-perception changes behaviour.
2. Self-perception can be manipulated using rewards, clothing, or virtual representations.

3. In order for self-perception to be changed, a symbolic meaning to the manipulation must be available to the person.

4. Stereotypes contain sufficient symbolic meaning to shift self-perception.

5. If a virtual representation has health-related symbolism, then health-related behaviour may be altered.

Following this line of reasoning, inferences may be made about the effect of avatars with different body types on in-game exercise behaviour. If items 1 to 3 are taken for granted, then manipulation should be possible from item 4 onwards. Given the apparent reverence of athleticism, and stigmatisation of weight in many cultures, manipulating the body type of an avatar ought to be sufficient to evoke a shift in self-perception. Given that the Proteus Effect relies on stereotypes, the behavioural assimilation should vary in the ways predicted by other social psychological theories, such as impression formation and valence asymmetry.

2.8 Research Question Revisited

This literature review has demonstrated that increased physical activity is beneficial for health, and that early attempts at influencing active behaviour through exergames may have been positive. It has also been shown that avatars may have behavioural modification properties, which sparked attempts to inspire players to exercise at a higher intensity by eschewing avatars with larger bodies in exergames. However, problems with limiting representation through the narrowing of avatar affordances were described. Further, the potentially deleterious effects of exposing people to depictions of 'fit-others' in the real world were outlined. Finally, gaps in the literature regarding the internal construction of athletic bodies, and implications of these in the Proteus Effect were identified. Based on this scoping review, the research question was reframed in the following way:

What are the implications of playing a game with an athletic, vs a plus-sized avatar, and how might these affect health-related behaviours in a virtual environment?.

The methods that will be used to address this question will be detailed in Chapter 3. In brief, the programme consists of exploratory and confirmatory phases. A systematic review was conducted to locate and evaluate existing theory, methods, and findings on avatar appearance as a behaviour change method across all health-related behaviours. Qualitative methods were employed to explore the experience of playing an exergame with an athletic, self-similar avatar, and to contrast this with gamers personal avatar creation habits. To address the gap in the literature on general athletic stereotypes, exploratory and confirmatory studies were conducted to address the structure these representations, and the extent to which these can be manipulated using additional behavioural information. In the final study, an attempt was made to link these stereotypes with behaviour through a replication and extension of a previous laboratory study (Li et al., 2014).

Chapter 3

Methods

> The first principle is that you must not fool yourself - and you are the easiest person to fool
>
> Feynman (1974)

This Chapter will outline the general methodological approach to the PhD, as well as the key methods used in the studies. The mixed methods approach was adopted (Johnson & Onwuegbuzie, 2004), meaning that the thesis is comprised of qualitative and quantitative studies. A description of this approach, and justifications for using it are provided in § 3.1 (page 62). Next, The Author's ontological and epistemological assumptions will be outlined, with more complete accounts provided in the appendices (https://osf.io/jhra6/).

Where possible (and/or meaningful), a Bayesian statistical framework has been used to analyse the data from the quantitative studies. The justification for using this will be provided in § 3.8 (page 92). For readers who are unfamiliar with Bayesian statistics (the majority of statistical training focuses on the so called *Frequentist* paradigm), § 3.8 provides a brief primer on Bayesian statistics, which may safely be skipped by those who have prior knowledge[1].

[1] Although §3.8 (page 91) provides the general approach that I took for selecting prior distributions and may be of interest to all readers.

Many of the studies in this thesis were pre-registered on the Open Science Framework, and the materials and data are available to download and run. The reasoning behind this decision is presented in § 3.3.

In addition, the methods that were used throughout the study will be described and evaluated. First, the use of avatars and exergames in previous research will be discussed, as well as the justification for the game that was used as a basis for the interview in Chapter 5. In Chapter 4, a systematic review was pre-registered and conducted following the PRISMA protocol guidelines (Moher et al., 2015). In Chapter 5, an interview schedule was developed, and the data were analysed using Thematic Analysis (Braun & Clarke, 2006), and a discussion of these methods will be presented in §3.5. In Chapters 6-8 the stereotypes associated with fitness and plus-sized individuals were assessed using descriptive, static and dynamic visual stimuli. The development of these stimuli is covered in Chapter 6. The analysis methods, such as Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Bayesian Regression, will also be described from § 3.6 (page 78).

3.1 Mixed Methods Approach

A mixed methods research programme involves the use of both qualitative and quantitative methods. Given the breadth of the research aims, a choice was made to run the programme in exploratory and confirmatory phases. The exploratory phase involved predominantly qualitative methods, including a systematic review (Chapter 4), and interview study (Chapter 5), but also an ideographic exploration of the stereotypical representations associated with different body types (Chapter 6). The confirmatory phase involved hypothesis testing using pre-specified models (Chapters 7 - 9). Johnson & Onwuegbuzie (2004) argue that methodological *pluralism* or *eclecticism* result in “superior” research when compared with single methods (i.e. *monomethodology*). They review two types of design, ‘mixed-model’ and ‘mixed-method designs’, and contrast these with the common monomethodological approaches in the sciences. The authors argue against the so-called *’incompatibility hypothesis’*, that quantitative and qualitative approaches are mutually exclusive.

Mixed Model and Mixed Method Designs

In the mixed model approach described by Johnson & Onwuegbuzie (2004), methods are blended together, and both may feature in any one study. That is, the data may be analysed from quantitative and qualitative perspectives. The mixed methods approach involves phasing of approaches, such that there is a qualitative stage and a quantitative stage (or vice versa). Johnson & Onwuegbuzie (2004) distinguish between *across stage*, in which a qualitative element might be added to a quantitative design (or vice versa); and *between stage* designs, in which methods vary between studies. The phases of mixed methods designs are demarcated, and methods will fluctuate between these phases. That is, there are qualitative and a quantitative phase, with the one method informing the later phases of the study. The current thesis can be seen in terms of qualitative and quantitative phases, with the qualitative methods being used to develop research questions which are examined in the quantitative studies.

3.2 Perspectives and Meta-theory

In the quantitative studies, a realist ontology with an empirical epistemology was adopted. During the qualitative interpretive study of Chapter 5, parts of the Critical Realist ontology were used, and a form of phenomenological epistemology was adopted (Willig, 2013). However, the underlying assumptions were that the structures that the qualitative methodology converged upon could ultimately be observed and measured. Critical Realism is not as appropriate for experimental work[2] since Critical Realism relies on abductive reasoning, settling on the "best possible answer" (Chirkov & Anderson, 2018). Further, since a portion of the thesis aims to critically address previous theories and methods, Lakatosian defence will be briefly reviewed as the main critical philosophy of science underlying the thesis[3].

Ontologies

The following will be a brief discussion of ontologies in general, and the different ontological perspectives. Ultimately, this thesis adopts a realist, and in places borrows from a Critical Realist perspective. An ontological structure is also discussed in Chapter 8 (page 256) to describe individual versus group level stereotypes, and this section will be referred back to.

An ontology describes the properties, relationships, and interactions between classes. A class in an abstraction, that, when instantiated, creates an object with attributes inherited from the class. These attributes include properties, interactions, and categories. For instance, the class *Feline* has the properties 'four legs', 'tail', 'whiskers', 'meows', 'purrs'. There are categories of felines (big cats, Maine Coons[4], Burmese, moggies, etc). Members of the same class *Feline* have the interactions within the class, 'hisses at', 'cuddles up with', 'steals food from'. The class also has interactions with other classes, e.g. the interaction with class *Rodent* is 'chase, kill, present as a gift', and with class *Human-Slave* 'wake up at 5am because the food bowl is not completely full'. When an object of class feline is instantiated, (e.g. the cat, Zigmund Farrow-Manfred, Figure 3.1), it will inherit these properties, relationships, and interactions[5].

[2] Although Chirkov & Anderson (2018) argue that much can be gained from a Critical Realist approach to quantitative research

[3] For brevity, the following is a summary of a larger document. Interested readers may find the full subsection on the OSF Appendix page https://osf.io/k7jrn/.

[4] Forest cats, hunters.

[5] ZFM does indeed pester for food, hisses, has four legs. However, he is an old boy with few teeth, and sadly his rodent killing days are long behind him.

Figure 3.1: The cat Zigmund Farrow Manfred being a loaf

Scientists, philosophers, and other theorists subscribe to ontologies of reality in their work. An ontology of reality describes the hierarchy of classes of objects that exist, their properties, categories, and interactions with other objects. An object may be physical matter (particles, books, person-as-object), but also latent abstracts such as society, media, and family. Two major (realism, relativism), and one minor (Critical Realism) ontological perspectives discussed below.

Realism and Relativism

A realist approach to the structure of knowledge assumes a stable reality in which objects are assumed to be independent of human consciousness (i.e. there exists no relation between the two). This is the most common ontology within 'pure' scientific domains (e.g. physics, chemistry). This approach assumes that it is possible to observe, quantify, and measure phenomena, and that repeated measurements *ceterus parabus* (all else being equal), will yield the same answer. Many subfields in Psychology attempt to be realist, i.e. they assume that behaviours, ideas, thoughts, etc are 'real', as well as more abstract functions such as neural networks (Dienes, 2008).

Alternatively, relativism supposes that one true reality does not exist, or rather that no single conception of truth is greater than another. Although relativism allows for tolerance for cultural, religious, identity purposes, extreme relativism becomes problematic when applied to the quantitative scientific method because consensus about objects is required (Dienes, 2008)[6].

Critical Realism

Critical Realists assume that there are underlying powers and structures in the world that exist independently of experience (Patomäki & Wight, 2000). These exist in addition to states of affairs, experiences, and accounts which are generated by the underlying powers. According to Bhaskar (2013), these exist hierarchically on two levels of reality, whereas empirical realism treats these levels as collapsed into one. Under a Critical Realist approach, there is a strong emphasis on the creativity of the researcher to explain reality based on 'empirical regularities' (Bhaskar, 2013), and by collecting observations of these regularities, the underlying mechanisms may be inferred. Purposive sampling is a feature of critical realist research - the sample that is most relevant to the theory should be used (this was the case in Chapter 5 but not the subsequent studies).

As a caveat, although Critical Realism offers some useful and interesting insights, and where such an approach may offer flexibility in the analysis of qualitative research, it is important to note that the thesis itself is not Critical Realist. The *structures* identified in Chapter 5 were subsequently tested as latent variables in the latter Chapters (Chapters 6 to 9), that is there is an assumption that these will be indirectly measurable, malleable through manipulation, and relatively static[7]. As such, the thesis takes a predominantly realist perspective, but with element borrowed from Critical Realism for interpreting the qualitative data.

[6] Indeed, by allowing for multiple truths, the 'Alternative Facts' statements made by Kellyanne Conway in 2017 suddenly become less absurd.

[7] i.e. the process of malleability is constant, as is the default state of the latent variable

Epistemology

Where ontologies describe the structure of knowledge, epistemology is the field of philosophy that is concerned with how knowledge is acquired, and what it is to *know* something. Different stances are taken throughout this thesis depending on the individual research question (each empirical chapter has at least one). In Chapters 6 to 9 there was a heavy assumption (even a necessity) that behaviour, perceptions, and attitudes may be measured and manipulated, and as such are approached from an empirical, primarily nomothetic approach. Empiricism assumes that only experience can produce knowledge, in contrast to the rationalism of Descartes et al who reason that knowledge may be rationalised from within (Kail, 2019)

The qualitative elements were partly informed by a phenomenological perspective. The basic unit of evidence for phenomenologists is not something that can be measured directly. Knowledge is gained through observing *capta*, or units of conscious experience. Objectivity is of little interest in phenomenology, and a staunch relativism is assumed (Willig, 2013). As such, the approach in Chapter 5 is somewhat a bastardisation of phenomenology,[8] and the accounts of participants experiences are assumed to represent something that exists in the *realist real*[9].

Lakatosian Defence

Under the meta-theory of Lakatosian defence, as a theory develops it gradually forms a 'hard core' of evidence, which becomes surrounded by a metaphorical protective bubble ("protective belt"), which requires increasingly stronger

[8] Since the staunchness of my relativism is minimal.

[9] This is reflected partly in the approach taken in Theme 4 (page 169) where during the analysis patterns in the language were observed which provided insight into how phenomena like weight stigma manifest, warranting an approach that could be described as an exploration of how the Bhaskarian *hidden powers* are experienced.

contrary evidence to burst, toppling the theory[10]. According to Lakatos (1979), theories ought to be judged on the extent to which they generate new insights. A healthy theory will continue to develop and make new predictions which may be corroborated by evidence (although this is not a necessary condition); such a theory is said to be on a *progressive* research programme. An unhealthy theory is one that relies on *ad hockery* to protect it from criticism, of which Lakatos describes three varieties: *Ad* hoc_1 is when theory that has no new content over competing theories[11]; *Ad* hoc_2 is a theory that can produce predictions but cannot corroborate them; *Ad* hoc_3 is a theory that is *patched* up with disconnected ideas on the fly[12]. Such a theory is said to be on a *degenerative* research programme or engaging in *strategic retreat* and can be conceptualised as being in a row boat that has several leaks in it. The rower spends more time attending to the leaks than rowing the boat to their intended destination. Under repeated failures to replicate or generate new theories, there comes a point at which there is no protective belt remaining, and theory can no-longer reject contradictory evidence. At this point, the hard core of evidence is affected, and the truth-value of the theory ought to be rejected. Lakatosian meta-theory encourages researchers to be critical of their criticism. No single observation is sufficient to topple a well formed and supported theory, as in Popper's Falsificationism which advocates for hard tests of a theory; instead emphasis is placed on forming a well-structured research programme that rigorously tests a theory. When replicating a study, or using a theory to generate new inferences, it is important to consider this.

Applied to the current research programme, the *hard-core* is the Proteus Effect. There are also a variety of auxiliary theories that are associated with this theory, including impression generalisation, structure of stereotypes, counter-stereotypical information. The hard-core has a wide variety of support,

[10]For many years, psychology existed in an economic bubble, in which theories were supported by poor investments. It was perhaps inevitable that this bubble would burst, leaving a circle of soap residue on the surface (see § 3.3).

[11]The Proteus Effect offers content over the competing explanation of simple priming.

[12]e.g. Yee's explanation of the height-negotiation effect changed when the original effect failed to replicate, see page 39.

which is presented in a general meta-analysis (Ratan et al., 2019), and for health-related behaviours specifically in Chapter 4. Given the general effect, and the auxiliary theories, new insights were generated and tested. This meta-theory will underly the critical interpretation of the empirical studies within this thesis.

Digital Dualism

An important final note on philosophy pertains to the concept of the *digital realm*. Proponents of this notion are broadly named *Digital Dualists*, and this standpoint involves the belief that digital activities are distinct from those in the *physical realm*[13], and have equal status to the realm of the *real*. Under digital dualism, a player is capable of “discarding the meat” of their body (Kolko, 1999), and *becoming* different identities[14]. This is a controversial standpoint, and one that The Author does not subscribe to[15]. The utopic, highly post-modern, digital dualist view of separate realms and identities has been criticised for disregarding the pre avatar-embodiment experience - which is, after all, the engine that drives the avatar (Kolko, 1999). For instance, a text-based role-playing game is practically useless to one who has no experience using written language to communicate! For the purposes of this thesis, the assumption is that digital activities exist within the *real*, and are a product of human action, rather than being independent of the physical.

[13]“Meat Space” as Andrew Przybylski named it in his 2019 Royal Society presentation.

[14]It all sounds very grandiose, and there are some rather creative descriptions of this process. For instance, Stone (1991) describes the process of entering the digital realm as disembodiment followed by “reembodi[ment] in the polychrome, hypersurfaced cyborg character of the console cowboy”.

[15]Indeed, The Authors view is that humans are highly skilled in extending their body using tools. Playing a game as a bi-pedal cat-like creature (e.g. Skyrim’s *Khajiit*) will make the user no more a feline than using a tool to clear leaves away makes the user a rake.

3.3 Open Science and Reproducibility

Since a realist ontology has been assumed in the quantitative aspects of this thesis, care is needed to avoid the impact of hidden flexibility. That is, where a single reality (universalism) is assumed, it is possible to bias the interpretation of this reality using various researcher degrees of freedom. It is through these degrees of freedom that the psychologist Daryl Bem managed to *prove* the existence of extra sensory perception (Bem, 2011). This was a achieved using a variety of (by the standards of the time) entirely legal moves such as optional stopping (continuing to collect data in the event of a non-significant effect); unplanned comparisons; additional covariates; stimuli splitting etc[16]. While psychology was coming to terms with the fundamental laws of physics being incorrect[17], a further impossible "finding" was published in the same year. In a *reductio ad absurdum*, Simmons, Nelson, & Simonsohn (2011) demonstrated that, using similar statistical methods to Bem, participants were chronologically younger than they were prior to hearing the Beatles song *When I'm Sixty-Four*[18]. Next, a large-scale multi-lab collaboration demonstrated that many *golden goose* studies may in-fact have been false positives ("Estimating the reproducibility of psychological science," 2015). Having broken Psychology, a movement called the *Open Science* reform, or *credibility revolution* began. This involved suggestions that researchers share their data, materials, and analyses; pre-register their methods and hypotheses; and focus on collaborations (Chambers, 2017). Where possible, this PhD follows open science best practices. All data and novel materials have been made available on the Open Science Framework (OSF), analyses were conducted using free, open source software (R, R Core Team (2017)), and a number of the studies were pre-registered (Chapters 4, 7-9). The thesis itself is a self-contained, reproducible

[16]In the interest of full disclosure, I myself have used all of these methods in my previous academic life between 2004-2010. Fortunately, only one copy of my master's thesis exists, and is on a dusty shelf somewhere at the University of Hull.

[17]The physicists were not worried.

[18]A *reduction ad absurdum* since in this exercise the assumption that $p < 0.05$ is an indicator of truth is taken as a given, and allows impossible hypotheses to be supported.

manuscript, meaning that when the document is compiled, all analyses may be run (it will take about a day to complete, so I'd suggest using the object files instead), the in-text descriptive statistics are populated, and graphs and tables are generated and formatted. This aids transparency, accuracy, and allows interested users to see the origin of each result. This was achieved using RMarkdown (Allaire et al., 2018), the Prepare APA Journal Articles package (Aust & Barth, 2017) and Bookdown packages (Xie, 2016). Each paper written from the thesis is, or will be, available as a pre-print from PsyArXiv (DOIs provided at the start of each chapter).

3.4 Sampling

In Chapter 2 (page 14) , the target population of people in the UK, between the ages of 16-44, who might enjoy videogames was discussed. Throughout this thesis, a variety of sources were used for data collection, and these were quite heterogenous. In Chapter 5, and Chapters 6, and 9, participants were predominantly students at Manchester Metropolitan University. In Chapters 7 and 8, they were respondents on a crowd sourcing platform. Figure 3.2 shows a *very* basic model of the assumed hierarchical structure of the population from which the studies will sample. It presents gamers as nested within society, and exergamers as nested within the population of gamers. By society, I mean, at its most specific, the population of the UK, and at its most general, the population of Western Educated Industrialised Rich Democratic (WEIRD), which have been the focus in the majority of avatar related exergaming research (see Chapter 4, page 128).

Under these assumptions, gamers *listen* to general society-level influences, such as stereotypes regarding different body types, since this was a feature of the environment we assume *the average gamer* was raised in. However, we do not assume that society *listens* to the influences of gamers - that is there are

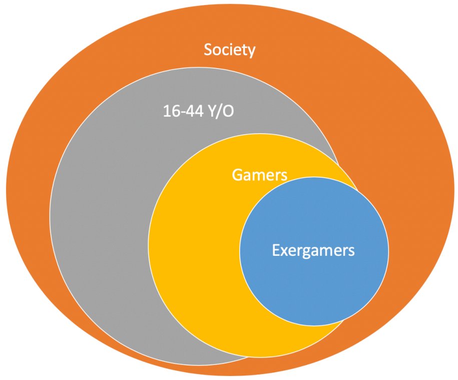


Figure 3.2: Population assumptions. This is for illustrative purposes and is not to scale. The portion of yellow that falls outside of the grey represents the consideration that some exergamers might not be gamers.

attributes of gamers that are specific to them, and not the larger parent group. Given this model, we assume that the monolithic *society*, the sub-population *16 to 44-year olds*, gamers, and exergamers hold the attribute *knows stereotypes*. Within gamers, a group also play exergames, and these people will have the attributes of gamers, but also be interested in exercising with their console. Not all gamers will be exergamers, but we assume that most exergamers are also gamers[19].

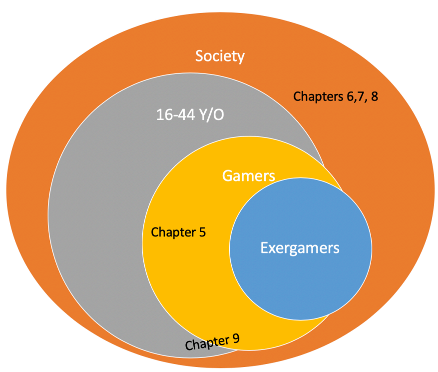


Figure 3.3: Population assumptions. This is for illustrative purposes and is not to scale. The portion of yellow that falls outside of the grey represents the consideration that some exergamers might not be gamers.

Under this model, I assume that attributes at the society level are

[19]Arguably some exergamers might not identify as gamers, e.g. casual, or *fad* players. The portion of the exergamer circle that excludes gamers represents this unknown quantity.

represented within gamer and exergamer populations, and so inferences made from sampling at a more general population may also be applied to the nested members. Conversely, we do not assume that the attribute 'has avatar preferences' is held by the whole of society.

I have no strong theoretical grounds for believing that knowledge of social stereotype-existence at the gamer level would be different to that of the general society, since these are inherited from the wider society. This is related to *In Here - Out There* stereotype distinction. The former refers to personal beliefs, which would be expected to change between sub-groups and the wider population, whereas *Out There* stereotypes are drawn from social and cultural norms and would not be expected to vary between groups (Kowert et al., 2012). However, I do acknowledge that the content of videogames differs somewhat in the degree of idealisation to more traditional television viewing (Martins, Williams, Harrison, & Ratan, 2009; Martins, Williams, Ratan, & Harrison, 2011), and this may influence stereotype content and availability. However, I make the (unverified) assumption that the gaming content is additive to the representations: that is, average gamers will be exposed to videogames, as well as television and social media which are ubiquitous in many societies, and will therefore have access to similar knowledge as the non-gaming members of the population[20]

An important limitation of the sampling plan in the studies of this thesis is that it will not allow generalisation outside of the UK. Both laboratory studies in this thesis were conducted at a UK University. It was not feasible to extend data collection to other countries. Moreover, the online survey studies that were conducted, although in theory are available for people in other countries to participate in, were limited not only to English speaking people, but also required an internet connection. In a reply to a comment by Gosling et al (2010) suggesting that the internet may be a bridge for non-WEIRD denizens to

[20]The alternative would be an interaction effect whereby exposure to videogames would attenuate or diminish the effect of general media influence. I have yet to see any evidence that this may be the case.

participate in Psychological research, Henrich, Heine, & Norenzayan (2010) note that global internet penetration (as of 2009) was low and unequally distributed across continents. For example, where 75% of people in North America had internet access, only 19% had access in Asia, and 6% in Africa. Moreover, people who have access to the internet in non-WEIRD countries are more likely to share characteristics with people from WEIRD countries, for example being educated or rich.

In short, I argue that general beliefs can be sampled from either sub-groups (gamers), or wider populations, but population-of-interest variables (e.g. gamers between 16-40), such as responses to avatar appearance or experience of exergaming, need to be sampled from that specific group.

Figure 3.3 shows how the different populations map onto the different chapters.

3.5 Exploratory Phase - Qualitative Methods

Study 1 Systematic Review

Study 1 involved evaluating the existing literature on avatar appearance-based health persuasion. The aim was to conduct a meta-analysis on existing effect sizes, however too few studies met these criteria for this to be meaningful. Instead, a qualitative assessment of the existing studies was undertaken. This consisted of both an interpretation of studies based on quantitative data (reported effects), and on societal values that appeared to be represented in the papers. Systematic Reviews are a robust method for synthesising research on a subject and are deemed to be the 'reference standard' for such tasks, due to the rigor involved (Moher et al., 2015). They follow a protocol that is ideally designed prior to data collection, and pre-registered. The protocol will contain information about search terms, sources, inclusion and exclusion criteria, methods of synthesis, and quality controls (Moher et al., 2015). The general aim of a Systematic Review is to (as far as reasonably applicable) comprehensively summarise a research topic and may include meta-analysis (quantitatively summarising a set of effects), or meta-synthesis (qualitative summarising of findings).

Systematic reviews are not without criticism: Ioannidis (2016) highlights the overuse of the method to 'mass produce' poor quality and 'redundant' reviews[21]. He states that only a small proportion of reviews are correct and useful, but the majority are, at best, correct but non-informative (e.g. simply highlighting that evidence is weak), and at worst wrong and misleading. Finally, he argues that resources would be better spent improving primary research prior to the use of secondary research.

A systematic review was deemed necessary for the current PhD to identify gaps in the literature, generate hypotheses from existing theory (given the weight of evidence across studies), and to provide informed priors for future Bayesian analysis. The protocol was written following the Preferred Reported Items for Systematic review and Meta-Analysis guidelines (PRISMA), and pre-registered on the Open Science Framework (https://osf.io/nzwpy/). The PRISMA guidelines are a 17-item list, presented in Figure 3.4.

The full PRISMA statement may be found at the OSF link above. It includes a detailed description of the rationale, methodology, intended analysis plan in the event of a meta-analysis, and details of qualitative synthesis in the event that a meta-analysis is not possible.

The difference between the pre-registration document and the final systematic review is to be noted and will be discussed in greater detail on page 131. The initial aim was to quantitatively synthesise the evidence using meta-analysis methods. The quality and quantity of available studies was overestimated, and a detailed plan was devised for analysing a large number of effects and investigating moderation effects. Further, the initial inclusion criteria

[21] In gaming and software, such products are referred to as *shovelware*.

Table 3 PRISMA-P 2015 checklist: recommended items to include in a systematic review protocol[a]

Section/topic	Item #	Checklist item
ADMINISTRATIVE INFORMATION		
Title		
Identification	1a	Identify the report as a protocol of a systematic review
Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	If registered, provide the name of the registry (e.g., PROSPERO) and registration number
Authors		
Contact	3a	Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support		
Sources	5a	Indicate sources of financial or other support for the review
Sponsor	5b	Provide name for the review funder and/or sponsor
Role of sponsor/ funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION		
Rationale	6	Describe the rationale for the review in the context of what is already known
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
METHODS		
Eligibility criteria	8	Specify the study characteristics (e.g., PICO, study design, setting, time frame) and report characteristics (e.g., years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	Describe all intended information sources (e.g., electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated
Study records		
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review
Selection process	11b	State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (i.e., screening, eligibility, and inclusion in meta-analysis)
Data collection process	11c	Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	List and define all variables for which data will be sought (e.g., PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data		
Synthesis	15a	Describe criteria under which study data will be quantitatively synthesized
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g., I^2, Kendall's tau)
	15c	Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression)
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned

Figure 3.4: PRISMA check list items (Creative Commons BY license)

included *studies that somehow altered an avatar*, and studies that *measured an outcome before and after exposure*. No studies met these criteria, and so the opportunity was taken to qualitatively synthesise the data and identify weaknesses in existing methodology to resolve in later studies.

Systematic Reviews and Meta-analyses are common in videogame research. There are several related to exercise games and video-game interventions to combat health issues (Guy, Ratzki-Leewing, & Gwadry-Sridhar, 2011; Lamboglia et al., 2013; LeBlanc et al., 2013; Matallaoui et al., 2017; Peng et al., 2012). In relation to avatars, there are fewer reviews. Fox et al. (2015) ran a meta-analysis on the persuasive power of human controlled avatars vs computer controlled 'agents', finding that participants were generally more influenced by human controlled representations. Ratan et al. (2019) present a meta-analysis of Proteus Effect studies, providing an aggregated estimate for both behavioural, and attitudinal outcomes. There have also been a number of narrative reviews of the potential applications of avatars and virtual environments in health settings

(Ahn, 2015b; Ahn & Fox, 2017b, 2017a). However, the use of avatars in health has not been directly and systematically summarised.

Risk of Bias

There are a variety of tools available when addressing the risk of bias in previous literature. The method employed in Chapter 4 was the Cochrane Collaborative Tool for Assessing the Risk of Bias. This is predominantly aimed at clinical studies, with a heavy emphasis on blinding, randomisation, and transparency. At present there is no consensus on assessing risk of bias in non-RCT clinical trials, and the Cochrane tool is possibly too conservative since it assigns high risk of bias for items that are not typically reported in Psychology studies, however clinical trials are a gold standard for experimental research, and studies that adopt their methodology may be more robust than those that do not. The Cochrane tool has been used by previous systematic reviews of psychological constructs, for example, Freijy & Kothe (2013) applied the method to Cognitive Dissonance interventions in a variety of health behaviours and produced useful information and advice for future research.

Study 2 Interview

To explore the experience of playing a videogame with a virtual representation of the self, an interview study was run in which participants discussed their general gaming and health behaviours and habits, had an avatar created that was based on their physical appearance, played an exercise game, and were interviewed about the whole experience. The interviews were analysed using the Thematic Analysis methodology as described by Braun & Clarke (2006). A variety of possibilities were considered when approaching the analysis of the qualitative data. These included the methodologies of Discourse Analysis and Interpretative Phenomenological Analysis, which come pre-packaged with assumptions about the acquisition of truth (Braun & Clarke, 2006; Willig, 2013). Although both participants' experiences, and evidence of latent constructs hidden in the patterns of language were of interest, The Author therefore decided to employ the method of Thematic Analysis, which affords researchers the flexibility to apply their own assumptions (Braun & Clarke, 2006). More details about this will be provided in Chapter 5. The majority of questions asked to participants were open-ended, such that they could not be answered with "yes" or "no". Some probing questions could be answered with closed answers (e.g. "do you, or have you ever used customisable avatars?"), but these were always followed up with probing questions. The interview schedule was developed using a funnelling method (Tengler & Jablin, 1983). Earlier questions were broad, and gradually narrowed down to specific topics. The order of topics was also set so they became more personal as the interviews progressed. Since interviewees were predominantly self-identified gamer, the interview schedule began with questions on their gaming history to help form a rapport. Further open-ended questions were then asked about the participants general health, followed by their satisfaction with their bodies. During the gaming session, participants were encouraged to talk about their experience of the game and voice their opinions about it in real time. This is broadly in line with an unstructured interview (King, Horrocks, & Brooks, 2018).

3.6 Quantitative Phase

Since the aim of the project was to identify reliable changes in performance as a function of avatar-appearance, the majority of the work within this PhD is quantitative. Although a qualitative approach could have been taken, the decision was made to address issues with the quantitative approach to avatar persuasion through rigorous experimentation. Several quantitative approaches were taken throughout the thesis: the frequency of idiographic responses to stimuli (Chapter 6); exploratory factor analysis of word ratings (Chapters 6 - 8); confirmatory factor analysis of factor structures between stimuli (Chapters 7 & 8); multi-level regression of word ratings (Chapters 6 - 9); regression of course completion times (Chapter 9). Each study in Chapters 6-9 were experimental in design, in that participants were randomly assigned to a condition; however only Chapter 9 was a laboratory-based experimental study. Working from a quantitative perspective, it is important to identify ones ontological and epistemic approach, which in this case is realist-empiricist, and to some extent positivist. That is, there exist objects with properties and relations, and these may be identified through measurement.

Latent Variables

Latent variables are hypothetical constructs that are assumed to have distributional properties and, from a realist perspective, to cause behavioural responses. A salient example is general intelligence. This is not directly observable like height, weight, or hair colour; but must be inferred through observations of behaviour, typically using a battery of intelligence tests. This thesis also handles latent variables as a major outcome. Chapters 6, 7, 8, and to some extent 9 all involve unobserved, inferred constructs - namely assumed qualities of stereotyped groups. This state of affairs is more complicated than a simple IQ scale since there are several moving parts which will be described in § 3.6.

Latent Variables and Realism

Since latent variables are merely hypothetical constructs that cannot be directly observed, the questions is raised as to whether realism can account for them. Certainly, there is space[22] within Critical Realism for the unseen, but these are searched through accounts and interpretation through collections of *empirical regularities*.

Borsboom (2005) however suggests that traditional realism can and does

[22]Indeed, a requirement!

account for the unobservable. He presents two path diagrams, one reflective and the other formative (Figure 3.5). In the *formative model*, (left panel of Figure 3.5) the individual items make up (i.e. *form*) the latent variable, meaning that item covariances must be modelled separately. Borsboom argues that this formulation is more compatible with a constructivist approach than a realist approach. The formative model is more of a summary of attributes, built up from multiple observations. For example, socio-economic is a latent construct that is socially constructed, and may be directly measurable through measures such as salary. A change at the item level affects the latent variable of SES. The individual items may correlate, but each can affect the latent variable independently[23].

In the reflective model (right panel of Figure 3.5), the latent variable has primacy over the items (the items *reflect* the latent variable) under the reflective model. That is, the items are assumed to be caused by the latent theory, and as a result the covariances are explained by the common root. The latter is primarily used by Psychologists, for instance the Big Five personality model assumes that Introversion causes ratings on the items related to the introversion sub-scale. There is an implicit assumption that a *Big Five* exist, and these cause the scale responses i.e. they are observable, ergo compatible with realist assumptions[24].

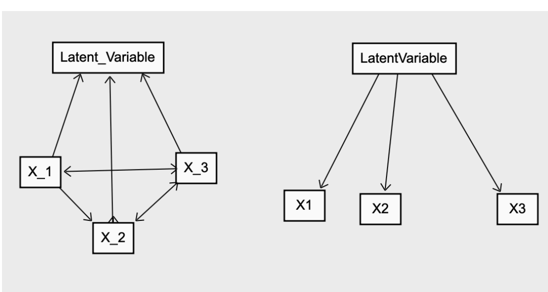


Figure 3.5: Approximation of the two path diagrams in Borsboom, 2005 p61

Borsboom (2005) argues that because primacy is given to the latent variable under the reflective model, they may be regarded as a realist entity.

[23]As a fairly ineffective example, receiving a scholarship to attend a private school may affect SES independently of parental income.

[24]To paraphrase Hacking, if it can be measured, and predict behaviour, it must exist

This conceptualisation also makes sense from a Qualitative perspective. The assumption was that the accounts around the body were influenced by media portrayals and social standards; as these standards change it can be expected that the accounts will change too. Interviewing at a particular moment aims to capture the state of latent structures such as an 'ideal body'. Although the ideal differs for each person, the values are assumed to be drawn from social norms and this would be expected to change over time.

Latent Variables in the Thesis

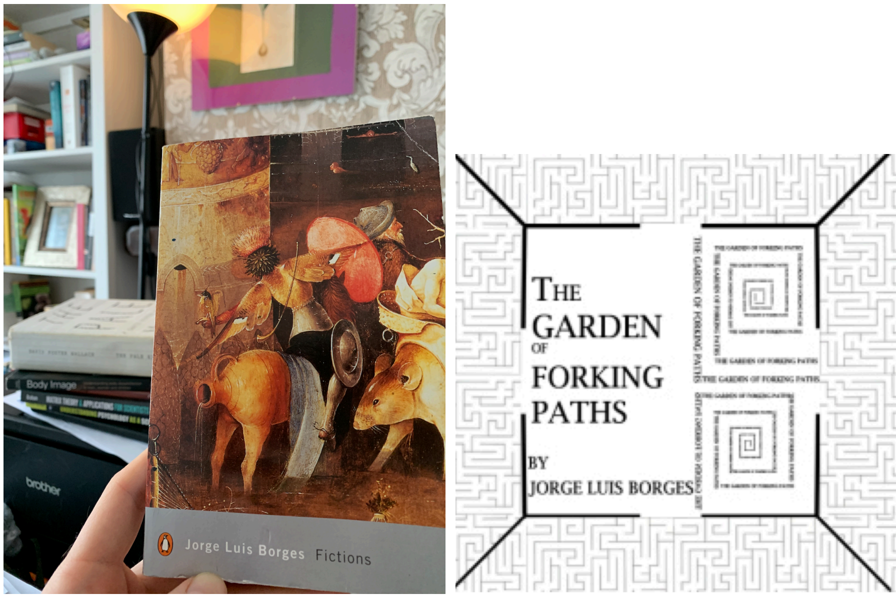
The exemplars used in these Chapters (be it verbal, visual, or animated) are assumed to fall on one of the formal $StereotypedAttribute_i$ constructs[25]. The point on which the exemplar sits is associated with a range of probabilities that a word generated from the formal construct (the set of which is the operational construct) will be rated on one of the $k-1$ scale points (e.g. Strongly Agree to Strongly Disagree)[26]. These probabilities are an indicator of the assumed behaviour or traits, marginalised over all subgroups represented in the sample, of one of the exemplars; however it is also an indicator of the actual behaviour of the participant, each of whom has an independent probability that they will select one of the $k-1$ thresholds in response to an exemplar. In addition, these probabilities may vary as a function of individual differences, for instance age, gender, body-type, body satisfaction. Modelling these factors would result in an exponential number of forking paths[27]. Individual variation of ratings was accounted for as far as possible by including random intercepts in each of the

[25]It will later be argued that the exemplars may be used as anchors for the extreme ends of each distribution, and that the space between these two anchors may contains less extreme body types - although this was not investigated

[26]More detail about this is provided in § 3.6

[27]A metaphor used by Andrew Gelman & Loken (2013), taken from the Borges story 'The Garden of Forking Paths' (Borges & Hurley, 1998) and adopted by the scientific community to represent the inconceivably large number of decisions that may be made when conducting analysis; For example, Orben & Przybylski (2019) calculated that the number of possible routes through the garden was in the order of trillions in their multiverse analysis of screen time and adolescent mental health.

models[28], but group-membership analysis (i.e. the physique of the participant) did not form part of the pre-registered analysis plans. This will be addressed further in the future directions of Chapter 10 (page 356).



(a) Cover of the Garden of Forking Paths

(b) Fictions by JL Borges

Figure 3.6: The second Borges reference in this thesis (although the first was a reference of a reference, within a reference). The Garden of Forking paths is a short story that touches upon the idea of a book that contains every possible combination of events, alluding to multiple parallel universes.

Level of Data

The majority of the quantitative measurements in this PhD are on an ordinal scale, i.e. Likert data. The reasoning behind Likert Scales is that they are a granular representation of a continuous, but unobservable latent variable (Likert, 1932). Likert Scales are often treated as metric (i.e. continuous), with the probabilities of each point of the scale being equal to the density at the corresponding point on the continuous normal distribution. This is inadequate because a maximum Likert response (of say 5) not only captures the point value 5, but all values above

[28] i.e. each participant is not assumed to have a common intercept, rather this is expected to vary.

5 (5 - ∞) on the continuous scale (Liddell & Kruschke, 2018). In a same way, the Likert value 1 captures all values up to one from $-\inf$. Further, the psychological distance between 1 and 2 may be different from 4 and 5, that is the intervals may not be equal between values - a major assumption of metric-level data. Moreover, it is difficult to know where the scale is centred on this hypothetical distribution, i.e. where the cut-points are. As demonstrated by Liddell & Kruschke (2018), the assumption that ordinal data may be treated as metric can lead to any of the following outcomes: a good estimate of the effect; a false positive effect; a false negative effect; a reversal of the effect. This clearly creates problems when interpreting data, since literally any outcome is possible; see Figure 3.7 for the diagrams provided by Liddell & Kruschke (2018).

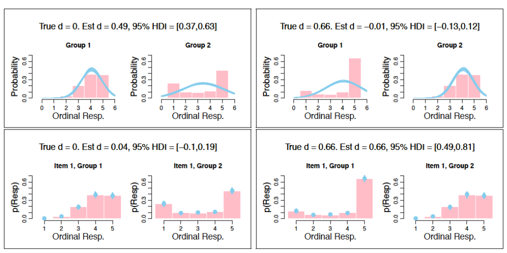


Figure 3.7: Taken from Liddle and Kruschke (2018): Ordinal data from a single item for two groups, displayed in histograms (same data in upper and lower panels). The left column shows a false alarm (Type I error) by the metric model (corresponding to points A and B in Figure 4). The right column shows a miss (Type II error) by the metric model (corresponding to points B and D in Figure 4). Posterior predicted data probabilities are shown as in the upper panels has a smattering of normal curves and in the lower panels as dots. The metric model (normal curves) are a poor description of the data, but the ordered-probit model accurately describes the data probabilities. This figure is licensed under CreativeCommons0 and may be accessed at https://osf.io/9h3et/.

Suggestions have been made to instead treat Likert data as a categorical, ordered representation of a latent continuous variable (Figure 3.7 bottom two panels). Fixing Likert data to an ascending order means that statements about the probability of a value being selected may be made. The laws of probability state that all probabilities must sum to 1. If 10 out of 100 people select a 1 on a Likert Scale, we know that there is a 90% chance that a value greater than 1 may be selected. Likewise, if 40/100 people select the numbers 1-3, we know that in this sample $P|y > 3 = 0.6$. Applying the inferred cumulative probabilities from a sample to a normal distribution function provides the approximate position of the Likert Scale points (cut-points) on a continuous normal distribution. On this 5-point scale, only the position of 4 cut points (in general $k - 1$) need to be estimated. This is because $k - 1$ cut points will segment a line (the range of z values on a standard normal distribution) into k chunks.

Analysing Likert data at an ordinal level therefore involves using a Generalised Linear Model, rather than the widely used General Linear Model (Bürkner & Vuorre, 2018; Liddell & Kruschke, 2018). In a GLM, the parameters for a distribution of a discrete outcome (e.g. counts, binary categorical, cumulative frequencies) are modelled, rather than the outcomes themselves. For instance, a count may only take positive integer values - but the single rate parameter (λ) may take any positive values (integer or float), and therefore may be modelled in terms of a linear model. Prior to estimation, the outcomes are transformed into the cumulative probabilities of the $k - 1$ scale points being selected to estimate the position of the points on a standard normal distribution. The cumulative probability is then modelled as a linear function of a set of predictors, which provide an estimate of the shift in probability of a cut point increasing from k_i to k_{i+1}. The predictor slope (β) is applied equally to all cut points in the case of cumulative logistic ordinal regression[29]. In brief, an ordinal regression table will have estimates for $k - 1$ intercepts (cut points), and an estimate for each predictor β_i and interactions. The β slopes represent the estimated 'mean difference', in z scores, of the latent outcome distribution (DV) per unit change of the predictor (IV)[30].

Ordered data also poses a problem for methods of determining latent

[29] Although it is possible to measure the individual effects of the slopes on cut-points - see Bürkner & Vuorre (2018).

[30] The reader will be spared the technical details, but Liddell & Kruschke (2018) and Bürkner & Vuorre (2018) are both excellent papers for outlining the issue with, and solution to treating ordinal data as metric.

variables such as factor analysis, which is commonly reliant on Pearson's Product-Moment Correlation (r). Treating metric data as ordinal has similar problems with accuracy in this context. These issues include over-dimensionalising of factor structures (Van der Eijk & Rose, 2015)[31]. Van der Eijk & Rose (2015) also note that even when data are treated as categorical, there is still a danger of overestimating the number of dimensions in a factor analysis, and provide a list of suggestions for minimising this risk, of which several were adopted in this thesis (use of parallel analysis rather than EFA, using polychoric correlations). These recommendations are also supported elsewhere by psychometrics authors (Flora & Flake, 2017; Holgado–Tello, Chacón–Moscoso, Barbero–Garcia, & Vila–Abad, 2010). A similar approach of modelling cumulative distributions is taken when estimating polychoric correlations (Drasgow, 2004). A polychoric correlation is the estimated correlation between two continuous latent normal distributions when data is only available as ordered categorical responses[32]. In cases throughout the thesis where data were correlated (including factor analysis and parallel analysis), polychoric correlations were used throughout).

The current PhD aimed to implement these recommendations where possible. In Chapters 6-8, Likert scales were used as dependent variables. Further, in these Chapters, Factor Analysis was used to estimate latent constructs that represent the stereotypes of athletic and larger-bodied individuals.

Further, ordinal regressions were used when analysing the extent to which these factors were affected by different representations of the physiques (discursive, visual, animated). In Chapter 9, interval data were collected in the form of time estimates and accelerometer data, and so the standard General Linear Model was used.

[31] An early attempt at analysing the data from study 6 resulted in the overdimensionalisation of the factor structure. When data were treated as categorical rather than continuous, the number of factors reduced from 5 to 3.

[32] For mixtures of dichotomous/polychotomous and continuous variables, point-biserial and polyserial correlations may be used. For technical details on polychoric correlations see Drasgow (2004).

3.7 Randomisation

Randomisation is important when running experimental studies. Selection bias can have large effects on estimates, negating or reversing observed effects. Throughout this project efforts were made to randomise and ensure blinding of the researcher to participants' condition assignment. For the most part this was handled using the *Qualtrics* block randomisation function because most of the quantitative studies were conducted online. This automatically assigns participants to conditions with no input from the researcher. In the exergaming study of Chapter 9, the order of participants was pre-specified using a script written in R which shuffled the conditions and created a two CSV files - one for men and one for women. The experimental software then queried the csv file with a participant index provided by the researcher.

Randomisation protects against two sources of bias: Selection bias and heterogenous treatment effect bias. In selection bias, one might knowingly assign a participant to a condition based on an *a priori* assumption that they will respond favourably or unfavourably to a condition. To bias the study in Chapter 9, an athletic participant could be assigned an athletic avatar based on the assumption that they would respond better to an avatar that reflected their physique (Cunningham, 2018). Heterogenous treatment effect bias is the bias that each participant will react differently to a treatment. For instance, it might also be that there is more variability in how participants with athletic bodies respond to their avatars, and so over-representing one group in the athletic condition may have its own heterogeneity. For instance, a participant may perform averagely with an athletic avatar where they would perform above their average with a plus sized avatar. This heterogeneity might not be present in people with larger bodies. By engaging in random sampling, rather than being purposive, both selection and heterogeneous treatment effect biases are mathematically negated.

3.8 Bayesian Statistics

A Bayesian approach was chosen for the modelling of quantitative data[33]. Classical statistical methods typically used in Psychology fall under the broad category of 'Frequentism' and are ideal for researchers concerned with the hypothetical long run frequency of an event occurring. Long run frequency refers to the set of all possible samples of a particular size that could be taken from the population (the sampling distribution). The general approach is to estimate a parameter from a sample that represents a theory (e.g. the difference between two means); if in the hypothetical long run, it is unlikely that at least as large a value would be observed under the null hypothesis of no effect[34], then the effect is deemed to be significant (Kruschke & Liddell, 2017). Since each sample is an independent observation of a parameter, repeated sampling contains no information about previous sampling, and so these methods are effectively amnesic[35].

Bayesian methods, on the other hand, take into account prior knowledge of a subject (belief), which after over 100 years of research in the field of psychology is likely to be available (false-positives not withstanding). Indeed, in Chapter 4, five studies were identified that contained relevant information about the effects under investigation, and in a recent meta-analysis into the Proteus Effect, a set of effect sizes were reported (Ratan et al., 2019). Bayesian methods allow both prior knowledge, and uncertainty to be embedded into the statistical analysis. The inclusion of *priors* creates a closed mathematical system from which probability statements may be made (Andrews & Baguley, 2013; Kruschke & Liddell, 2017). The Bayes-Laplace theorem[36] (equation 3.1) may be used to

[33]The following is marked IYI. It is a brief discussion of Bayesian statistics, which are used throughout this thesis. For a more thorough (and entertaining) contribution, see Etz & Vandekerckhove (2018).

[34]typically zero, although methods such as equivalence testing allow to test against a band of effects.

[35]Although previous literature may be used to inform sampling plans and smallest effect sizes of interest

[36]The theorem is commonly known as Bayes Theorem after Thomas Bayes, however Laplace

update prior knowledge with observed data to obtain a posterior-belief about an effect:

$$P(\theta|\mathbf{D}) = P(\theta)\frac{P(\mathbf{D}|\theta)}{P(\mathbf{D})} \tag{3.1}$$

Equation 3.1 is Bayes formula in algebraic form, but in English it simply states that *the probability of a hypothesis* (θ) *given the data* (D) *is equal to the probability of the data given the theory, multiplied by the prior probability of the theory, divided by the probability of the data.* The first term of the numerator is the observed data, the second term of the numerator is the prior belief in the hypothesis, and the denominator is the *prior predictive probability*, which is a weighted average probability of observing the data under all hypotheses in the model[37]. Where Bayes statistics becomes non-trivial is when data are continuous rather than single probabilities (Etz & Vandekerckhove, 2018). This requires the use of integrals to calculate the prior predictive probabilities, of which a finite number may be calculated analytically. This was an issue for many years, but with increasing computing power and a dedicated community of programmers and statisticians it became trivial to estimate prior predictive distributions using *Markov Chain Monte-Carlo* methods. The technical details of MCMC estimation are beyond the scope of this thesis, but an excellent discussion may be found in Chapter 8 of Richard McElreath's book *Statistical Rethinking*, (McElreath, 2016b). In brief, the probability density assigned to each posterior value is estimated using thousands of samples of random numbers. This process is represented visually in Figure 3.8[38].

---

did much of the work in operationalising the theorem and applying it outside of theology (McGrayne, 2011). Laplace's name is used here once to mark this fact, but hereafter will be referred to a Bayes Theorem or Bayesianism.

[37] In the case of alternative and null hypotheses, this is the sum of the probability (P) of the hypothesis (H_1) $\times$ probability (P) of the data (D) given H_1 $\times$ $(P|DH_1)$ and the $P|H_0 \times P|DH_0)$

[38] The colour scheme was chosen to represent a location from Stephen King's Dark Tower Series. The grey is the White Lands of Empathica, the red is the field of roses (Can'-Ka No

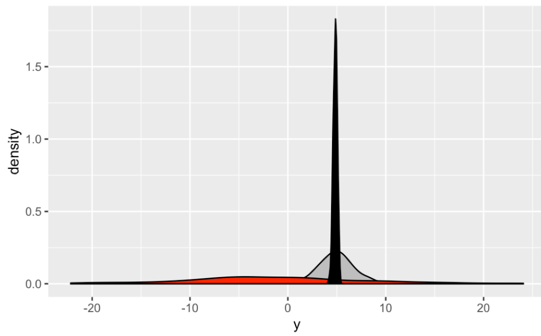


Figure 3.8: A graphical depiction of Bayesian updating. The red distribution represents a weakly informative prior over the range of values between ± 20. The grey distribution represents the spread of the observed data, and the black is the updated distribution which has shifted the probability density from the really extreme values and assigned more to the mode of the data.

Priors

One of the major issues that Frequentists have with Bayesian statistics is the necessity of priors, who commonly regard them as subjective and open to abuse. Indeed, this was one of the main reasons that Bayesian statistics were ostensibly vilified for much of the 20th Century (McGrayne, 2011)[39]. However, priors are merely the belief that we have in a hypothesis before observing the data. As Alex Etz points out, researchers must have a prior belief in a hypothesis, or they would not be conducting the experiment (Etz & Vandekerckhove, 2018). Probabilities are assigned to different parameter values in the form of distributions. In a uniform distribution, each value is assigned equal probability. There are very few situations in which a uniform distribution would be appropriate[40]. In most fields, some information is available that can inform priors, even if this simply regularises the posterior (i.e. constrains values to possible values). For instance, it is a fact that the reaction time of a human less than or equal to 0 is impossible[41], a reaction

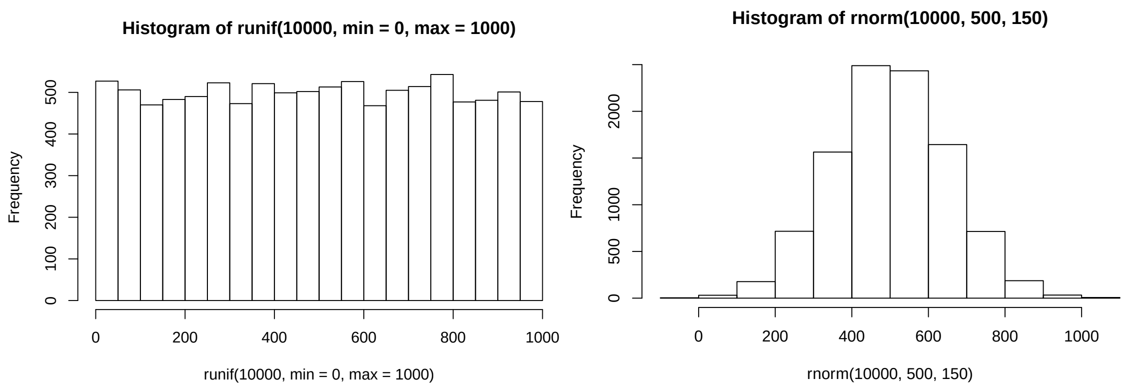
Rey), and the black is the titular Dark Tower (King, 2006).

[39] As discussed by Sharon Bertsch McGrayne in her book “The theory that would not die”, despite successes of Bayesian methods in applied settings such as cracking the Enigma in World War 2, it was deemed to be pseudoscientific by opponents such as Ronald Fisher.

[40] Although with sufficiently strong data and a well-defined model, a uniform prior may yield similar results to a frequentist interpretation

[41] Daryl Bem may disagree, since his participants reportedly had negative reaction times.

of time of 40ms is highly improbable, and reactions time above 10^{e5}ms is highly unlikely in healthy populations. Therefore, a uniform distribution ranging from $-\infty$ to $+\infty$ is beyond uninformative. Even limiting the range to 0-1000ms (Figure 3.9a) assigns as much probability to reaction times of 1ms as it does to 1000ms.



(a) Histogram of 10k samples from a uniform prior distribution between 0 and 1000. Each value is equally likely to be drawn

(b) Histogram of 10k samples from a normal prior distribution with a mean of 500ms, and a standard deviation of 150ms

Figure 3.9: Example prior distributions

Instead, we can centre a distribution on our best guess, and define uncertainty around it. We might, given previous literature, decide that 500ms is a fair estimate of the mean reaction time, but not be particularly certain. Choosing a normal distribution centred on 500ms with a standard deviation of 150 assigns a small amount of probability to extreme low and high values, with the majority (68%) of the density assigned to 500ms $\pm$ 150ms (Figure 3.9b).

With more observations, we become more confident with the parameters we are estimating (Royall, 2017). Indeed, the main function of Bayes-Laplace theorem is to update the prior distribution with what is observed. If the data are strongly in favour of the prior, then the posterior distribution will strongly resemble the prior. If the data are strongly against the prior, then the posterior distribution will shift, proportionally to the strength of evidence against the prior.

Selecting Priors

The general approach for selecting priors is discussed below. This approach was repeated in each study that used priors, and the decisions will be discussed in the relevant chapters.

In selecting priors for the projects in this PhD, similar studies were consulted to get an estimate of existing evidence. Based on the strength and quality of the evidence, random effect simulations were run based on the identified distributions, and likely ranges of values given these parameters were drawn from random number generators. Using samples from the posterior distribution, predicted simulated values, given the prior, and the idealised data were constructed using linear modelling, and plotted. This ‘prior predictive check’ identifies whether the prior produces impossible values when used to generate samples. If these values fall outside of the known probability space (e.g. 10,000 repetitions of an exercise), then the prior is fine-tuned until a reasonable maximum is drawn.

Interpreting Bayesian Regression Models

The output of a Bayesian regression model is much the same as a frequentist regression table. Terms such as intercept, a beta coefficient, and estimates of dispersion (standard deviations) are all comparable to frequentist parameters. However, instead of confidence intervals, which estimate the long run probability that false values will be excluded under repeated sampling, Bayesian models can provide *credible intervals* which are statements of the probability of a range of values, given the data and the prior (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). Because Bayesian models provide distributions as output, factual statements about the probability of the range of values may be made about the estimates. The credible interval gives the two points between which a portion of the probability density is assigned (e.g. a 95% credible interval is contains 95% of the probability density, and values outside of this are only 5% likely to be observed).

In addition, the number of effective samples used by the Monte-Carlo sampler to make these estimates is included in the table, as well as an estimate of convergence of the Monte-Carlo chains. Often trace plots are presented showing the agreement of the Monte-Carlo chains, and the overall health of the model.

Justification

The benefits of using Bayesian methods in Psychology are well documented (Etz & Vandekerckhove, 2018; Gelman, Hwang, & Vehtari, 2014; Kruschke & Liddell, 2017; Marsman et al., 2017; Wagenmakers et al., 2018), and they are becoming increasingly accessible and popular with the advent of high level wrappers[42], for example: Rstan, (Stan Development Team, 2018); brms, (Buerkner, 2016); rethinking, (McElreath, 2016a); and graphical user interface based software such as JASP (Team & others, 2018), and JAMOVI[43]. The decision to go full Bayesian was made because there exists a degree of prior information about the Proteus Effect, and some of this includes the application to health-related behaviours. Since the goal was to update what was already known about the Proteus Effect and body sizes, Bayesian methods were appropriate, especially since this information could be used to apply principled priors to the statistical models.

3.9 Replication Studies

A replication is an attempt to recreate the findings from a previous study using a new sample, whilst maintaining the original methods as far as possible (Bonett,

[42] A wrapper is a set of functions that reduces the number of lines of code required to perform a computer programming task.

[43] This is not to say The Author has any ill-will against p-values. There are many persuasive discussions advocating the use of these (Lakens, 2015; Daniel Lakens et al., 2018), however it was decided early on that Bayesian posteriors provided the information that was of most interest to The Author.

2012; Hendrick, 1990). Confidence in the replicability of an effect is essential in the quantitative sciences; Popper (1963) stated that any effect that is not replicatable must not be treated as scientific evidence, but merely as ‘chimeric’, and Dunlap maintained that repeatability is a necessary condition for science (Zwaan, Etz, Lucas, & Donnellan, 2018). Despite this, the publication of replication studies was relatively rare in Psychology until recently, accounting for a mere 1 percent of published studies in 2012 (Field, Hoekstra, Bringmann, & Ravenzwaaij, 2019; Makel, Plucker, & Hegarty, 2012). Replications may take many forms, and broadly fall along the close and far (e.g. similar sample to the original vs vastly different sample), and direct and conceptual (very similar method vs different method used) dimensions. Conceptual replications, in which different measures are used to address a theory, may also be described as ‘extension’ studies whereby the same theory is applied to a broader context. Under healthy scientific conditions, this suggests that a theory can make new predictions that may be corroborated and is a feature of a Lakatosian progressive research programme.

Field et al. (2019) suggest three considerations when deciding on which study to replicate: statistical, theoretical, and methodological[44]. First, results of previous studies should be reanalysed using Bayesian statistics to estimate the probability that the findings would occur under the alternative hypothesis [45] ; next the replication should be qualitatively assessed in terms of theory and should address a theoretically important and currently relevant effect. Finally, it should be possible to repeat the methods used, particularly if technical, and the skills of the researcher should be considered.

Although these recommendations had not been released as a paper at the

[44]Note that this article post-dates the beginning of the project, and indeed the decision to replicate a previous study.

[45]The authors suggest the use of Bayes factors which were not discussed in § 3.8, but essentially compare the likelihood of the data under the experimental hypothesis with the likelihood of the data under the null hypothesis, providing a ratio of likelihoods. A BF10 [46]of 8 means the data are 8 times more likely under the experimental hypothesis than the null hypothesis.

[46]This notation means the Bayes Factor of hypothesis 1 against hypothesis 0 (the null); BF values that tend towards infinity are in favour of H1, whereas those that tend towards 0 are in favour of H0. If the notation were BF01, the reverse would be true.

time of planning the thesis, a similar, but less refined thought process was behind the decision on which study to replicate. Previous research was assessed (not using Bayesian methods) for quality and similar studies were grouped and assessed for replicability (Chapter 4); the theory that seemed most appropriate for generating new inferences was chosen (i.e. Proteus Effect), and methods were assessed with the ability of the researcher in mind. In Chapter 9 one of the Proteus Effect studies described in Chapter 4 was replicated. Rather than choosing to replicate studies that used a competitive tennis game which would have required substantial games development skill, The Author decided on a simple running track game which replicated Li, Lwin & Jung (2014), who used a Wii running game. Were the project to be repeated, the recommendations of Field et al. (2019) would be followed.

3.10 Choice of Stimuli

Avatars in Exergame Research

The Nintendo Wii console was a pioneering console for avatars and exergames. It was the first mainstream videogame console to offer a customisable 3D character (a Mii) that could be designed to look like the player and was the first to successfully implement motion sensing technology[47]. With the release of the Wii in 2006, there was a surge in the popularity of avatars and exergames. Following from the Wii, Microsoft released their customisable avatar system in 2008, and Kinect camera for the Xbox 360 in 2010, and Sony released their PlayStation Move controllers which mimicked the Wii controllers, also in 2010. The accessibility and popularity of these features was followed by a wealth of research on the effects of both avatars and exergames.

[47] There were other products, such as the PlayStation EyeToy which were not as successful.

In Qualitative Research

Research adopting a Qualitative approach towards player-avatar interactions is relatively sparse. The majority of these studies use open-ended questionnaires (e.g. Cole & Grogan, 2019; Cacioli & Mussap, 2014; Ducheneaut, Wen, Yee, & Wadley, 2009; Kafai et al., 2007b; Kafai, Fields, & Cook, 2010), and few use a rigorous qualitative-psychological approach which is grounded in theory (e.g. Cole & Grogan, 2019). In other studies, an interview approach was used to investigating people and their avatars (e.g. Ford & Cole, 2019; Banks & Bowman, 2013; McArthur, 2018; Neustaedter & Fedorovskaya, 2009; Shaw, 2015). These interviews were conducted in a range of contexts, including face-to-face while playing a game (Shaw, 2015), or while designing an avatar (Ford & Cole, 2019; McArthur, 2018). In other interview studies, the interviews were conducted in an online virtual environment such as SecondLife (Neustaedter & Fedorovskaya, 2009) or World of Warcraft (Banks & Bowman, 2013; J. Banks & Bowman, 2016a). The interview techniques included semi-structured and unstructured interviews.

Banks & Bowman (2013) used unstructured interviews in the online setting of World of Warcraft (a Massively Multiplayer Online Role-Playing Game, or MMORPG). The unstructured approach involved meeting with a participant in the virtual environment of Azeroth and asking, “What shall we do today?”. The interviews developed during gameplay and included discussion of the player’s avatar. From her dataset she constructed a framework for understanding the relationships that participants have with their avatars, including avatars as objects, avatars as tools, avatars as symbiotes, and avatars as the self.

McArthur (2018) used a method called ‘micro-ethnography’ to investigate the decisions made by people when designing avatars, and reactions to having an avatar created for them using the WiiU camera. McArthur describes the micro-ethnography as a method of capturing small events and analysing them using an interactionist framework. The sessions were filmed so that specific instances, and the participants responses to them, could be returned to and analysed in depth. Ford & Cole (2019) interviewed participants on their avatar design choices during an avatar creation session. They report that the women in their study adopted prescribed gender norms regarding the body size of the avatar but rejected gender-role based stereotypes. Interviewing participants during gaming activities has many benefits. Participants can relay their decisions and reactions to events in real time, rather than reflecting upon the experience first. In Chapter 5, participants were also interviewed immediately before and after an avatar generation and gameplay session, so that the experience may be discussed after reflection.

In terms of exergaming, qualitative methods have been used to investigate the efficacy of game-based alternatives to exercise. Methods include focus groups (Wollersheim et al., 2010), observational studies (Finco, Reategui, Zaro, Sheehan, & Katz, 2015), and semi-structured interviews (Lam, Sit, & McManus, 2011). Qualitative exergame studies predominantly focus on the feasibility of exergames as a replacement for traditional forms of exercise. That is, there have been no in-depth qualitative psychological studies on the experience of exergames and the role that avatars play in that experience.

In Chapter 5, these two purposes are explored simultaneously. Of interest was the way in which both self-representations, and the embodied experience of videogame-mediated exercise interact and inspire motivation to exercise.

In Quantitative Research

Previous quantitative research into avatars aimed to investigate the interaction between player and avatar in terms of promoting presence (Ratan & Hasler, 2009), enjoyment (Trepte & Reinecke, 2010), identification (Li, Liau, & Khoo, 2013), health-related intentions (Jin, 2009),increasing body dissatisfaction (Sylvia, King, & Morse, 2014), and modifying behaviour (Yee & Bailenson, 2009; Yee et al., 2009). Avatars may either be customised by participants (e.g. Kim & Sundar, 2012; Jin, 2009), tailored to look similar to the participant (e.g. Song et al., 2013a), or designed to look a certain way by the researcher (e.g. Peña & Kim, 2014). In studies investigating avatar customisation, participants are generally instructed to create an avatar that resembles either their actual, ideal, or ought selves (Jin, 2009; Kim & Sundar, 2012; Sah, Ratan, Tsai, Peng, & Sarinopoulos, 2017). This is achieved through a variety of methods, including the online environment SecondLife, the Nintendo Wii *Mii* creator, and bespoke software.

The aims of quantitative research in the context of exergaming include determining whether it is as effective as traditional exercise (Peng et al., 2011), and whether it can influence people's attitudes, intentions, and future exercise behaviour (e.g. Van Nguyen et al., 2016; Li & Lwin, 2016). A typical exergame study involves participants playing an exergame either in isolation or in a group, and either having their performance measured or self-reporting on psychosocial measures.

In-game performance has been measured both by using accelerometers attached to wrists and waists (Peña et al., 2016; Peña & Kim, 2014), or by measuring the time taken to complete a level of a game (Li et al., 2014). The exergaming sessions can either be in isolation (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014) or provided over the course of several weeks (Li & Lwin, 2016; Lwin & Malik, 2012; Van Nguyen et al., 2016). Finally, a variety of studies have used Randomised Controlled Trial methods to investigate changes in physical ability after a period of exergaming (Huang et al., 2017; Rahmani & Boren, 2012).

The types of exergames used include dance-based games in which participants follow on screen instructions to perform dance moves, e.g. Just Dance 3 (Li & Lwin, 2016), sports games in which participants use body movements to mimic playing a sport e.g. Kinect Sports, Wii Sports (Li & Lwin, 2016; Peña et al., 2016; Peña & Kim, 2014; Song, Kim, & Lee, 2014), and personal trainers in which guidance is given to improve the players technique or general fitness e.g. EA Sports Active, Wii Fit (Van Nguyen et al., 2016).

In Chapter 9, participants were assigned an avatar that had been created by the researcher to represent groups of individuals with athletic and plus sized body types. An exergame was also developed using the avatars created in Chapter 6, using the Unity game engine, and motion-sensing controllers. The development of this software will be discussed in Chapter 9 (page 306). The game included double blinding of conditions, random assignment, and measured both the time taken to complete the course, and accelerometer data based on movement. The approaches discussed in the above paragraphs formed a large part of the planning of the studies in this programme. It was useful to critically appraise previous methods.

The Approach

In summary, the adopted ontology for this PhD thesis is Realism, with elements of Critical Realism for the qualitative study. However, it is assumed that both quantitative and qualitative methods are approaching the same, or at least hierarchically related constructs. A predominantly empirical epistemology is adopted throughout, with experience being the primary unit of evidence (either through measurement or interpretation of capta). The programme was run in several stages, over five empirical Chapters. These phases are summarised below:

1. **Determine** what is already known and believed. This involved assessing the literature about avatars and health-related behaviours (Chapter 4).

2. **Contrast** this with the reported experience of creating and playing exercise games with avatars (Chapter 5).

3. **Reframe** the broad research question outlined on page 1 in terms of 1 and 2.

4. **Generate** inferences from previous literature and create new hypotheses which can be tested using the existing framework.

5. **Test** the hypotheses from 4 using a ground up approach (Chapters 6 - 9).

Chapter 4

A systematic review of research on avatar appearance and health-related outcomes

4.1 Rationale

In this chapter, existing theories, methods, and studies were located and evaluated. The purpose of this was to guide the research programme and to narrow down and aid the operationalisation of the broad research question and objectives discussed on pages 1 and 5. The chapter also directly addresses the first objective on page 5, to assess the strength and quality of evidence for avatar appearance as a health-related behavioural intervention.

The reframing of the research question on page 58 stated that the implications of using an athletic, versus a plus-sized avatar were of interest. This chapter will provide a systematic justification for this reframing, placing it in the context of a variety of other possible research questions (e.g. relating to *ideal-self* avatar customisation), and argue that previous studies investigating the prescribed appearance of an avatar on health-related behaviours are based on the strongest theoretical foundations, relative to other research in this area. It will also demonstrate that, of the available outcomes, physical activity is the most sensitive to changes in avatar appearance.

4.2 Introduction

There have been numerous attempts to *gamify* health (Alahäivälä & Oinas-Kukkonen, 2016). This involves using game-based rewards to aid motivation and create healthy habits. These attempts may focus on acquisition of a skill, or health behaviour; focus on promotion of health information or advice; or creating an entertaining alternative to exercise, *exergamification* (Wattanasoontorn et al., 2013). The latter may involve the inclusion of motion sensing controllers as part of the game mechanism, bridging the gap between gaming and physical activity.[1]

When playing videogames, audiences are often given the flexibility to choose their virtual appearance, whether this is through choice of game, character, or through integrated customisation systems. The end result is an avatar, or representation of the self in a virtual environment (Nowak & Fox, 2018).

The physical appearance of an avatar has been shown to elicit behaviour change in virtual environments. For example, Yee et al. (2009) found that people controlling an "unattractive" avatar in a virtual environment disclosed less personal information and maintained a larger interpersonal distance than those controlling an attractive avatar. Choice of avatar may also temporarily alter attitudes towards the self. Fox et al. (2013) found that women embodying a "sexualised" avatar self-objectified more than those who embodied an "un-sexualised" avatar.

If physical appearance during a video-game (or other digitally-mediated experience) can alter attitudes and behaviours, then digital self-representation

[1]The Pre-print for this article may be found at https://psyarxiv.com/j3675/.

may have the potential to improve health outcomes. The current study aimed to review the literature on the impact of avatar appearance on health-related outcomes.

There are a number of potential pathways for avatar influence on attitudes or behaviours. A direct route involves manipulating an avatar's appearance and altering the user's self-perception. This results in the adoption of behaviours or attitudes that are congruent with the adopted self-image. This route is called the Proteus Effect (Yee et al., 2009). As a relevant running example, manipulating an avatar to look athletic may lead a user to feel that they are *fitter* than they currently are while controlling an avatar. This may result in a temporary improvement in exercise exertion during a videogame-based exercise (exergaming) session or increased post-play intentions to exercise.

An indirect route involves allowing players to customise their avatars to represent actual or ideal representations of themselves in game. The latter may offer an outlet for wish-fulfilment. As such, the presentation of the self as an ideal may reduce the perceived discrepancy between the actual-self, and ideal-self. This ideal self may be athletic and encourage the user to engage in more healthy lifestyles (Kim & Sundar, 2012).

A third, more passive route is related to traditional media theory. In this, the mere exposure to idealised forms may be sufficient to reduce a person's body satisfaction, or exercise behaviours. This has been demonstrated experimentally in both real life (Crawford & Eklund, 1994; Wasilenko et al., 2007), and in response to magazine images and other media (Hall, Baird, Gilbert, Miller, & Bixby, 2011; Homan, McHugh, Wells, Watson, & King, 2012). By this line of reasoning, using an athletic or idealised avatar would result in upward social comparisons, and negative feelings towards the self. However, this simple interpretation has been criticised for suggesting that audiences are passive, critically unengaged with media content, and evidence in favour is limited (Ferguson, 2013).

4.3 Existing evidence

The role of avatars in health-related attitudes and behaviours has not been systematically investigated. There are a variety of reviews and meta-analyses relating to exergamification, in which players use larger gestures involving their bodies to control a game (Biddiss & Irwin, 2010; LeBlanc et al., 2013; Matallaoui et al., 2017; Peng et al., 2012, 2011). This increases physical exertion, and has been shown to require a similar level of energy expenditure to mild to moderate exercise (Peng et al., 2011). Most exergames have an avatar that responds to the players movements, however the actual role of the avatar has not been included in these reviews. Indeed, there are very few exergaming studies that investigate the role of the interaction between the player and avatar (e.g. Li & Lwin, 2016; Li et al., 2013). The majority of exergaming studies generally favour the social, or competitive side of exergaming (Song et al., 2014, 2013b).

At the time of writing, there are two previous reviews of the literature on health and avatars: a chapter in an edited book on virtual healthcare (Ahn & Fox, 2017a), and an entry in the Oxford Encyclopaedia of Communications (Ahn & Fox, 2017b)[2]. While both offer interesting insights into the possible uses of avatars and agents (computer controlled characters) in healthcare, neither systematically review, or assess potential sources of heterogeneity or bias in the literature. Both articles state that there is a surprising dearth of literature on avatars and health behaviour, and the majority of the research reviewed pertains to agents, or doppelgängers (non-user controlled 3D representations that are created to look like the user), rather than avatars. The authors of these reviews highlight that utilising agents in delivering healthcare interventions may be more cost effective than avatars which must be controlled by humans, and therefore are constrained by human biology and staffing limits. These reviews suggest that there are two unique affordances that avatars and agents have over traditional

[2]The author notes that two very similar systematic reviews were published shortly before submission (Horne et al., 2019; Rheu, Jang, & Peng, 2019), but had no impact on the planning or writing procedure.

communication methods: time acceleration (being able to show a participant a future possible self); and interactivity (allowing users to have agency over the media content). Time acceleration has been used to show weight loss and weight gain in response to high calorie food intake (Fox et al., 2009b), soda consumption (Ahn, Fox, & Hahm, 2014), and exercise behaviour (Fox & Bailenson, 2009). In these studies, participants observed a 3D model that looked like them facially, gain and lose weight whilst performing a health-related behaviour. In a number of these studies, targeted health behaviours were affected by this experience.

However, it is notable that none of the studies reviewed in the above chapters are related to avatars by the standard definition, that is, they are object representations (i.e. representation as a third party), rather than subject representations (representation as self). Nor do the reviews discuss the quality of the evidence for avatars affecting health-related behaviours. As such, this review will be the first to address evidence for the appearance of avatars in health related behaviours.

4.4 Objectives

The objectives of this review were:

1. To determine the efficacy of avatar-appearance in improving health-related behaviour outcomes including direct observable behaviours; direct (e.g. intentions), and indirect (e.g. related to body dissatisfaction) self-reported measures.

2. To address the quality of the methodology within the studies using existing tools (e.g. the Cochrane tool for assessing bias, Statcheck, etc).

3. To use meta-analytical techniques to estimate the overall effect size for avatar health-related persuasion.

4.5 Method

The guidelines from the PRISMA statement were adhered to in the undertaking of this systematic review (Moher et al., 2015). The following is an abbreviated method section[3].

Search strategy

The studies for this review were located using electronic database searches of Medline, PsyArticles, PsychInfo, Pubmed, Web of Science, and Web of Knowledge. Unpublished papers were sought using the electronic thesis repository ProQuest, and by contacting prominent authors in the field. The search strategy was developed with the assistance of an academic librarian and the terms included *avatar, digital self, videogame, virtual reality, health, exercise, behaviour, change, attitude, or intention.* Searches were not restricted to English journals, although only languages that could be translated using Google Translate were considered. Reference lists, chapters, and author websites were also consulted.

Continuous Searching

Alerts were set for NCBI and Scopus. An email was received by the author whenever a study meeting the criteria of the search terms was released. Each paper was checked using the same three point method as described below (title, abstract, full text screening).

[3]Full details can be found in the digital appendix https://osf.io/k7jrn/.

Eligibility Criteria

Types of Studies

All quantitative studies looking at effects of avatar appearance on health-related outcomes, using either comparative or within-group methods were considered. This included conference papers, peer-reviewed articles, and unpublished dissertations.

Types of participants

Only studies investigating non-clinical populations were eligible for inclusion. Studies on children and adults were included.

Types of intervention

Only studies manipulating some element of avatar appearance were included in the review. Participants also had to have some control over the avatar. As such, studies in which the only manipulation was customisation, and those that had no avatar manipulation were excluded. Manipulations included customisation studies, and studies in which the avatar was assigned to a participant. When reviewing studies for inclusion, the definition of 'avatar' was identified as a major source of variation. The research question was consulted again, and the research team decided to define an avatar as a representation of the self in a virtual environment. Excluded from this definition were studies on virtual self-modelling, which rely on the participant seeing the model as different from themselves, or as a 'doppelgänger'. In these studies, participants typically observe the model, and have no control or influence over it. This was vastly different from the majority of studies selected for inclusion. VSM studies may warrant a separate systematic review. Similarly, virtual pet studies were also excluded on a similar basis, in that the avatar was not intended to represent the participant, but to act as a model that reflected their behavioural decisions such as eating breakfast (Ahn et al., 2015).

Types of outcome measures

Outcome measures included observed and self-reported changes in body-image and health-related attitudes and intentions. Observed differences in behaviours, such as eating or exercise, were also of interest.

Study selection

Studies from all sources were aggregated into a master-file and screened for title and abstract using Rayyan (Elmagarmid et al., 2014; Khabsa, Elmagarmid, Ilyas, Hammady, & Ouzzani, 2015). Initial screening was conducted by the author. Members of the supervisory team blind-reviewed 10% of the judgements made by the author. The remaining studies left after screening were subjected to full-text screening. Eligibility of ambiguous papers was discussed and resolved at research meetings.

Data Extraction

Data were extracted following the method of a previous literature review (Freijy & Kothe, 2013), and involved extracting the country, health outcome, sample characteristics, comparison group manipulations and relevant results. Inferential statistics for the relevant results were also extracted. The highest acceptable error rate to represent a meaningful effect for this review was 0.05, and so any studies reporting ‘marginal’, or ‘trends towards significance’ were interpreted as not significant.

Assessment of Risk of Bias

Risk of bias was assessed using the Cochrane Collaboration tool (Higgins et al., 2011). Each study was assessed for selection bias (randomisation assignment of participants, and concealment of allocation from personnel), performance bias (blinding participants to the intervention), detection bias (blinding of assessors and other personnel, attrition bias (loss of participants), and reporting bias (selective reporting of outcomes).

Data Synthesis

Due to the small number of studies, and degree of heterogeneity in outcomes and methods, the studies were not quantitatively synthesised. Rather, a qualitative approach to the synthesis of the data was adopted.

4.6 Results

Searches yielded 13 articles, of which one had two studies, making a total of 14 studies to be included in the review. Figure 4.1 shows the PRISMA flow diagram for study selection[4].

Post Hoc Study Categorisation

Upon reviewing the literature, the selected studies were categorised into two groups. These groups were not declared in the pre-registration, and so a brief discussion of this categorisation is warranted before discussing the results. A set of studies within the corpus involved providing participants with pre-made avatars with varying characteristics. These often varied by body-shape, but in one study [4] varied by attractiveness. This set of studies is labelled 'Avatar-Appearance

[4]Full details about the yields from each source can be found in the supplemental materials.

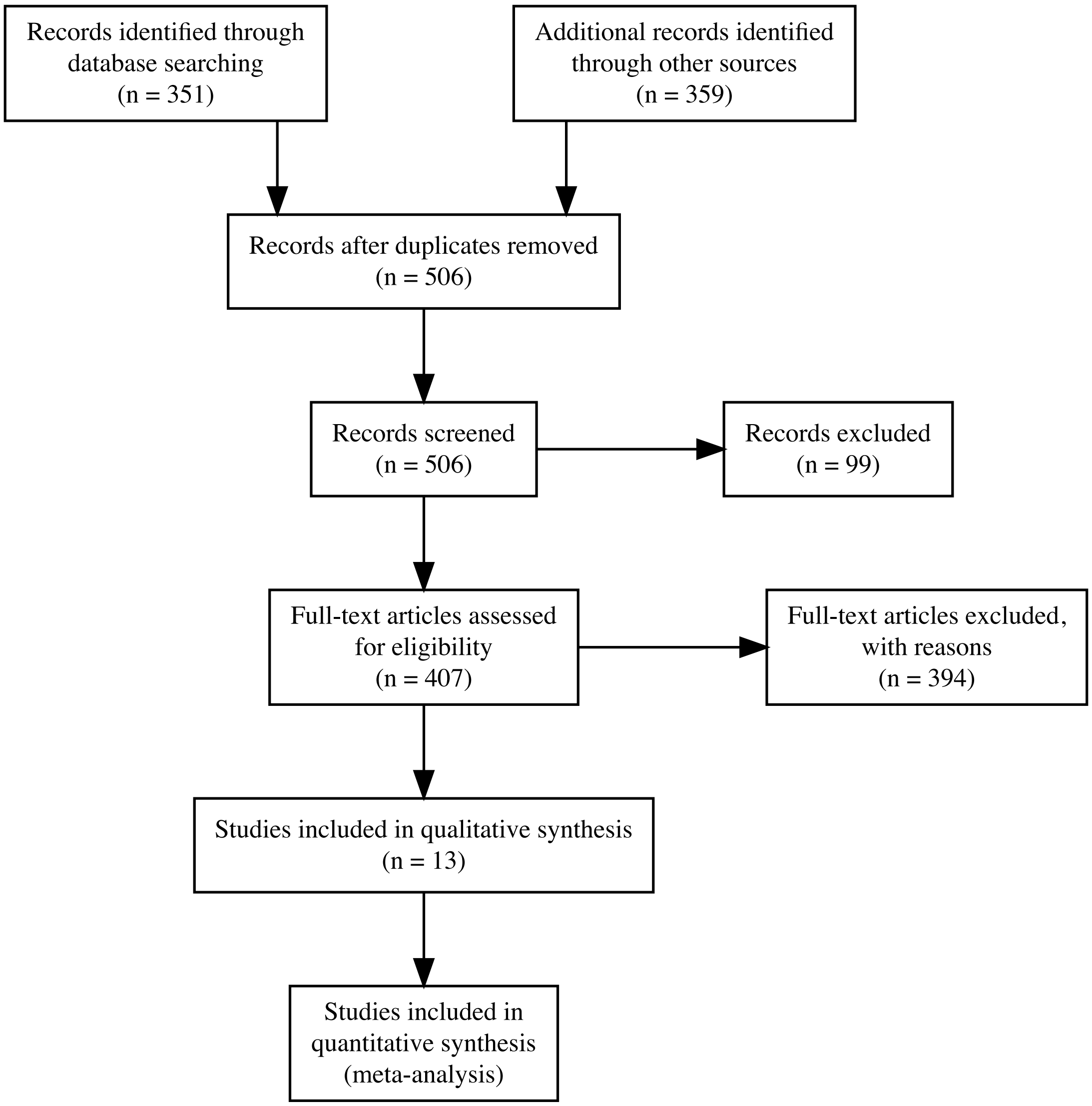


Figure 4.1: Prisma Flow chart

Studies'. The second set of studies involved asking participants to create an avatar that looked like themselves. Most provided instructions to create an avatar based on Self-Discrepancy Theory related self-models (Higgins, 1987) . Self-models are internal representations which consist of attributes that a person currently possesses ('actual-self'), attributes they would ideally like possess('ideal-self)', and attributes they believe they are morally obliged to have ('ought-self'). This second set of studies is labelled 'Avatar-Customisation Studies'.

Study Characteristics

Table 4.1 provides the study characteristics and identification numbers for the dataset. All studies were peer reviewed journal articles published in a range of journals including: 10 from human computer interaction journals [1,4,6,7,8,9,10,12,13,14]; three from psychology [2,5,10]; and one from a games and health journal [3]. One study had previously contributed towards a doctoral thesis [4].

Primary health behaviours included smoking [12], exercise [3,6,9,10], diet and [2,3,5,11,14], maintaining general healthy lifestyle [4], and Human Papillimamilliary Virus vaccination [1]. Secondary health attitudes included body satisfaction [7,8,13].

One study was conducted on children [6], selecting only overweight children. The remainder of the studies were sampled from undergraduate populations, with four studies selecting only males [1,8,10,13], and three selecting only females [7,9,11]. No studies reported sample size calculations based on statistical power analysis, although one reported an *a posteriori* power analysis [2].

Table 4.1: Summary of study characteristics

ID	Reference, country, and study design	Health-related outcome	Sample	Comparison group manipulations	Relevant results
1	Darville et al (2018) USA Randomised trial	HPV vaccination	N = 168 (108 completed) Full time students between 18-26 years	Participants either selected or created an ideal or actual self-avatar and played a serious game (Vax!) in which they had to stop the spread of a disease using vaccinations. Avatars were 2D images that remained on screen throughout the game.	No main effect of avatar type for perception of risk of HPV. No main effects of avatar condition for self-efficacy. No main effect of avatar type on behavioural intentions. Intervention itself was effective.
2	Jin (2012) USA Randomised trial	Diet	N = 156 Undergraduate students (103 women, 53 men), mean age 19.89 (SD 1.12)	Participants created a Nintendo Mii to represent their actual or ideal self, and played a Wii Fit exercise game (ski slalom, running and Hula-hoop).	No main effect of avatar appearance on low calorie dieting intentions.
3	Joo and Kim (2017) Korea Randomised trial	Exercise and eating	N = 124 Undergraduate students, Males = 45, females = 79, aged 20-29 years	Participants played The Sims with either an Obese or a "normal" avatar.	No main effect of avatar body shape on number of steps taken on a stepper. No main effect for avatar body shape on number of high sugar cookies consumed.
4	Kim and Sundar (2012) USA Randomised Trial	Healthy Lifestyle	N = 95 Undergraduate students (69 women, 26 men), aged 18-34 years	Participants either created an avatar to resemble their desired or actual self; or were assigned an attractive or unattractive avatar. They used the avatar in a virtual Centre for Disease Control in second life and interacted with an agent that explained the mission of the CDC, and demonstrated posters that included information about sleep disorders OR obesity.	Participants who created actual avatars had a higher 'general perceived risk to the physical body' than those who created desired avatars. Participants who customised their avatars said they would devote more time to maintaining good health than assigned avatars. People who customised their avatars selected a higher number of unhealthy product coupons than those who were assigned avatars.
5	Kuo et al (2016) Taiwan Randomised Trial	Diet	N = 76 Undergraduate students Females = 48, males = 28 Mean age = 21.2	Participants were immersed in a 3D virtual environment and presented with a fitting room. They were instructed to look in a virtual mirror and saw either their actual self or a weight reduced version. The ideal-weight self was tailored to the participants self-reported ideal size. They completed measures on temporal discounting and various filler measures before being asked to complete an ice-cream taste test and indicate the amount of sugar they would like in a reward Bubble Tea drink.	Participants in the weight-reduced self-condition ate less ice-cream during the taste test (once the last meal was controlled for) than those in the current self-condition. Participants in the weight reduced condition chose smaller amounts of sugar for the reward drink than those in the current-self condition. Participants in the weight reduced self-condition were more likely to choose a sugar free version of the drink than those in the current self-condition.
6	Li et al (2013) Singapore Factorial Randomised trial	Exercise	N = 140 Overweight students aged 9-12 years	Participants played a Wii exercise game with an avatar. Avatar had large or "normal" body size Stereotype threat induction or not	Main effects of avatar body size. Those embodying larger avatars had lower exercise attitudes, exercise motivation, exergame motivation, and game performance.
7	Matthews et al (2016a) USA Randomised trial	Body satisfaction	N = 149 Female undergraduate students aged 18-24 years.	Participants played video games either as an ideal (Beyond Good and Evil, Beijing Olympics 2008), or a hyper-ideal avatar (Mortal Kombat, Dead or Alive Beach Volleyball). There was a no-game control condition.	No difference in sexual Attractiveness, Higher (more favourable) scores on weight concern in hyper idealised condition than idealised condition; No difference Physical Condition; No difference in Body Discrepancy; No difference in Contour Bust Size Scale; Lower scores on Body Dissatisfaction sub-scale for participants in hyper idealised condition.

8	Matthews et al (2016b) USA Randomised trial	Body satisfaction	N = 197 Male undergraduate students, aged 18-30	Participants played video games either as an ideal (Alan Wake, Fight Night), or a hyper-ideal (UFC Undisputed, Street Fighter 4) avatar. There was also a no-game control condition.	No differences in physical attractiveness, No differences in upper body strength, No main effect on physical condition, No main effect on body discrepancy, No main effect on muscle Discrepancy, No main effect on positive attitudes towards muscularity.
9	Peña and Kim (2014) USA Randomised trial	Exercise	N = 96 Female undergraduate students, aged 19-25 years	Participants played a Virtua-Tennis game on the Wii. Avatar had ”normal” or ”obese” body.	Main effect of avatar body size. Women in normal body condition made more wrist movements than those in the obese condition. No effect on wrist movement.
10	Peña et al (2016) USA Randomised trial	Exercise	N = 96 Male undergraduates, aged 18-32 years	Participants played a Virtua-Tennis game on the Wii. Avatar had ”normal” or ”obese” body.	Main effect of avatar body size. Men in normal body condition made more wrist movements than those in the obese avatar condition. No effect of waist movement.
11	Sah et al (2016) USA Randomised trial	Eating	N = 133 Female undergraduate students Mean age 20.26 (SD = 1.34)	Participants created either an actual, ideal, or ought self in the bespoke virtual environment YooNot. They selected food items for the avatar in YooNot and made the avatar perform in-game exercise activities such as dancing and jumping traffic cones.	No difference between the actual and ideal group in the number of healthy food items selected by participants after the game. Fewer items selected by participants in the ought than the actual avatar condition.
12	Song et al (2013) USA Randomised trial	Smoking	N = 62 Undergraduate students, 38 males, 24 females, mean age 22.05 (SD 3.28)	Participants played a social smoking related serious videogame with an avatar that had their photo as a face, or another person photograph as a face. The face either had age related artefacts or not in the second level.	Multivariate differences between future face and not future face conditions, but not self and other avatar condition. No difference between future face groups on perceived risk, but sub group analysis showed there was a difference in the self-avatar condition; Participants in the future face group had stronger attitudes against social smoking than those in the no-future face group; participants in the future face group had stronger intentions to quit smoking. Participants in self condition reported stronger perceived susceptibility to smoking than those in the other condition. No differences between self and other avatars on any dependent variable.
13	Sylvia et al (2014) USA Randomised trial	Body satisfaction	N = 51 Male undergraduates, aged 18-24	Participants played Skyrim for 45 minutes. Avatar had either normal or muscular physique.	Differences in body satisfaction between hyper muscular and normal group. No differences in perceptions of muscularity.
14	Velhurst et al (2018) Japan Randomised trial	Health food shopping	N = 23 Students (21 male, 2 female), mean age 22.5, SD = 0.71.	Participants performed a food shopping task in an immersive virtual environment using either an obese avatar or a ”normal” avatar.	Participants in the obese conditions bought neither more healthy products nor less unhealthy products than those in the ”normal” condition. Participants in the obese condition neither perceived unhealthy items as healthier than those in the ’normal condition’.

Outcomes

Of the studies, six directly examined behaviour [3,5,6,9,10,11]: three aimed to promote physical exertion during the game [6,9,10]; one study looked at post-game exercise (number of steps taken on a stepper machine) [3]; three studies examined the number of food items consumed directly after play [3, 5, 11]. Attitudes were the primary outcome in 9 studies [1,4,6,7,8,12,13,14]. Behavioural intentions were measured in 6 studies [1,2,4,5,6,12]. In one study these were measured using a token-based lottery analogue involving coupons for healthy or unhealthy products or services; in a second they were measured using the amount of sugar to be included in a future drink. Only one studies measured attitudes, behaviours and intentions [5].

Behaviour

Behavioural outcomes included both in-game and post-game measures. In-game dependent variables included the number of healthy virtual food items selected by participants [11,14]. The degree of in-game performance was the dependent variable in three studies and was measured using accelerometers on the wrist and waist [9,10], and through timing how long it took for a participant to complete a running game [6] which was used as a proxy for how fast they were moving. Post-test influence outcomes included the number of steps taken on a stepping machine [3], the number of high sugar cookies consumed [3], amount of ice-cream eaten in a tasting-test [5], and the number of healthy snacks selected by participants [11]. No studies measured long term behaviour change.

Intentions

Intentions were measured using self-report methods that were either developed for the study or adapted from previous research. The development of these self-report measures was not discussed in detail, nor was a framework referenced for their development (Ajzen, 2002). Two studies attempted a direct measure of behavioural intentions using a coupon selection analogy [4] in which participants were invited to select vouchers for products or services that were either healthy or unhealthy; and through the amount of sugar to be included in a bubble tea reward drink [5]. The development and standardisation of these method is not discussed within the papers, and so the extent of their validity cannot be determined.

Attitudes

Attitudes were measured using a mixture of novel and standardised methods including body image measures such as the Swansea Muscularity Attitudes Questionnaire, Body Esteem Scale and Body perceptions scale. One other study measured attitudes towards health behaviours [6] using questions about participants opinions of exercise.

Reliability and Validity

For self-report measures, all but two studies [5,14] reported a measure of internal consistency (Cronbach's alpha, α). The reasons for selecting α rather than other measures of internal consistency (e.g. ω) was not provided by any study. Reliability of measurement ranged from 0.7 to 0.94. No studies discussed the validity of its measures. In some cases, standardised measures were used, and such information could, in principal, be acquired by following references. Table 4.2 provides details of the validity and reliability measures discussed in the papers.

Table 4.2: Summary of validity and reliability of measures

Reference	ID	Measure	Reliability	Validity
Darville et al (2018)	1	Adapted measures from previous HPV research.	Self Efficacy: $\alpha = 0.812$ Intention: $\alpha = 0.836$ Perceived Risk: $\alpha = 0.910$	Validity of behavioural intentions in predicting behaviour mentioned. Validity of measurement not discussed.
Jin (2012)	2	Novel measure: Development not described (Unsure if they used Ajzen's framework)	Low calorie dieting intention: $\alpha = 0.8$	NA
Joo and Kim (2017)	3	Behaviour	NA	Validity not discussed
Kim and Sundar (2012)	4	Novel measures and borrowed items	perceived Risk: $\alpha = 0.93$ Attitudes towards social smoking: $\alpha = .83$ Perceived susceptibility: $\alpha = .78$ Perceived Risk to Physical Appearance: $\alpha = 0.8$ Perceived Risk of general physical body: $\alpha = 0.7$ Self-perseveration: NR	Validity not discussed
Kuo et al (2016)	5	Observed quantity of ice-cream eaten and intended quantity of sugar consumed in the future.	No reliability reported.	Validity not discussed
Li et al (2013)	6	Novel measures	Exercise Attitude: $\alpha = 0.84$ Exercise Motivation: $\alpha = 0.9$ Exergame Motivation: $\alpha = 0.91$	Validity not discussed
Matthews et al (2016a)	7	Standardised and novel measures	Body Esteem Scale: SA $\alpha = .86$; WC $\alpha = .92$; PC $\alpha = .89$ Body Discrepancy: NR Contour Bust: NR Eating Disorder Inventory: BD $\alpha = .85$	Validity not discussed
Matthews et al (2016b)	8	Standardised and novel measures	Body Esteem Scale: PA $\alpha = .87$; UBS $\alpha = .91$; PC $\alpha = .87$ Swansea Muscularity Attitudes Questionnaire: PAM, $\alpha = 0.93$	Not clear whether accelerometer is a valid measure of exergame movement.
Peña and Kim (2014)	9	NA	NA	Validity of behavioural measure not discussed.
Peña et al (2016)	10	NA	NA	Validity of behavioural measure not discussed.
Sah et al (2016)	11	Novel measure	Health consciousness: $\alpha = 0.76$	Validity not discussed
Song et al (2013)	12	One modified measure, two novel measures	Perceived Risk: $\alpha = 0.93$ Attitudes towards social smoking: $\alpha = .83$ Perceived susceptibility: $\alpha = .78$	Validity not discussed
Sylvia et al (2014)	13	Standardised and modified measures	Swansea-Muscularity Scale: a= .916 Body Esteem Scale: $\alpha = .955$	Standardised methods used - validity not discussed.
Velhurst et al (2018)	14	Translated 'inspired' by peer reviewed research	NR	Validity of measures not discussed

Experimental Paradigms

Although there was a visual appearance element to the avatar manipulations in all of the studies, the task, type of variation, and type of avatar varied considerably between, and within studies. In studies where avatars were varied by body size [3,5,6,9,10,14], and muscularity [7,8,13], the studies were between subject designs in which the body type varied by group assignment. Customisation studies were far more varied. There were 9 separate manipulations, and few of these were repeated. The most widely used was Self Discrepancy Theory (SDT) (Higgins, 1987). The self-concepts and self-models from SDT were used as primes for the participants to create their avatars from (e.g. actual, ideal and ought selves).

Studies in which participants customised their avatars were more complex, with assignment and customisation being used as manipulations in some cases, and just customisation used in others. The customisation of avatars varied based on the instructions that participants were given, which broadly reflected the self-guides from Self Discrepancy Theory. They were instructed to create an avatar that reflected their actual, ideal, or ought selves in one study [11], and actual or ideal selves in three studies [1,2,4].

Evidence for effectiveness

Table 4.3 provides a summary of the significant effects from each study.

Most studies [N = 9] reported direct effects of avatars on health related attitudes and behaviours, with 7 reporting positive changes (an increase in health related attitudes, intentions or behaviours) [4,5,6,7,9,10,12], and five reporting negative effects [4,5,7,11,13]. Negative results were mixed across paradigms and outcomes. For manipulations involving obese avatars, performance was reduced, whereas manipulations involving idealised avatars reduced positive attitudes towards the self. Studies reporting negative effects were predominantly those manipulating avatar body type, with participants using larger ('obese')

Table 4.3: Summary of effects. ID = participants were presented with normative ideal avatars; SDT = participants were primed with Self Discrepancy Theory self-concepts when designing their avatars; OB = participants were assigned either an 'average' or 'obese' avatar; Future = participants were presented with a future version of themselves.

Health-related outcome	ID	Behaviour	Attitude	Intention	Manipulation
Physical activity	0	NS	.	.	OB
HPV vaccination	1	.	NS	NS	SDT
Diet	2	.	.	NS	SDT
Exercise and eating	3	NS	.	.	OB
Physical activity	4	+	+	.	CUST
Healthy Lifestyle	5	.	+	±	SDT
Diet	6	+	.	-	ID
Exercise	7	+	+	+	OB
Body satisfaction	8	.	±	.	ID
Body satisfaction	9	.	NS	.	ID
Physical Activity Self efficacy Anxiety	10	NS	NA	NS	SDT
Exercise	11	+	.	.	OB
Exercise	12	+	.	.	OB
Eating	13	-	.	.	SDT

bodied avatars reporting lower self-report outcomes or demonstrating a lower performance (conflated with physical activity) in game compared with those in the smaller ("normal") body type conditions. However, it is important to note that none of the studies reporting negative effects had an adequate control condition. Only one study [6] measured attitudes, intentions, and behaviours.

Assessment of Methodological Risk of Bias

Assessment of the risk of bias is presented in Table 4.4. The clearest message is that not enough steps were taken to control for (or report) bias in these studies. Only cursory mentions of random allocation were made, with no indication on the blinding of the manipulation or assignment from either participants or research personnel. Three studies [7,8,13] indicated that participants were allocated prior to arrival, meaning that selection bias could potentially be higher. One study described the method of randomisation as being a permutated block randomisation [5], but it was unclear when this process occurred. Three studies (two from the same author) included an awareness check to detect participants who had guessed the purpose of the study [5,9,10]. Attrition bias was not suitably addressed, with the majority of studies having discrepancies between the reported sample size and the degrees of freedom reported with their inferential statistics - and in two cases (from the same article) [8,9], the degrees of freedom varied between outcomes within the same study. The majority of studies used self-report measures, so observer bias was not applicable. Most studies reported all outcome variables, however in two, [9,10] covariates were omitted from the analysis due to the variable not being a significant predictor, and the values were not reported.

Table 4.4: Summary of assessment of risk of bias

ID	Reference	Adequate sequence generation	Allocation concealment	Blinding of participants	Blinding of study personnel/assessors	Incomplete outcome data addressed
1	Darville et al (2018)	?	?	?	?	?
2	Jin (2012)	?	?	?	?	+
3	Joo and Kim (2017)	?	?	?	?	-
4	Kim and Sundar (2012)	?	?	?	?	-
5	Kuo et al (2016)	?	-	+	-	+
6	Li et al (2013)	?	?	?	?	-
7	Matthews et al (2016a)	?	?	?	?	-
8	Matthews et al (2016b)	?	?	?	?	-
9	Peña and Kim (2014)	?	?	+	?	-
10	Peña et al (2016)	?	?	+	?	-
11	Sah et al (2016)	?	?	?	-	?
12	Song et al (2013)	?	?	+	?	-
13	Sylvia et al (2014)	?	?	+	?	+
14	Velhurst et al (2018)	?	?	?	?	-

Sources of Heterogeneity

There was considerable variation in how the studies were run, how the avatars were chosen or created, the activities that participants performed with the avatar, and the methods used to measure the outcomes. In 8 studies, the avatars were from consumer videogames [2,3,6,7,8,9,10,13]; in one study the 3D virtual world SecondLife was used for the customisation and presentation of avatars [4]; in 5 studies bespoke or specialist avatar creation software and environments were used [1,5,11,12,14]. In two studies, participants' photographs were used to create the avatar [5,12]. The main source of heterogeneity was whether an avatar was assigned to participants [N = 9], or whether it was customised by them [N = 4]. Two studies directly compared customisation with assignment [1,4].

The use of standardised measures did not appear to affect outcomes. Further, the inclusion of covariates was inconsistent throughout, with 8 studies reporting their use (two were removed for not being significant). Covariates included weekly game use, gender, BMI, body discrepancy, and body shape concern, and time since last meal. There was no mention of covariates in the hypotheses of the studies, so it is not clear whether these were *a priori* decisions.

The platform of the study was a further source of heterogeneity. The majority of studies were conducted using desktop computers or consumer console, but two used fully immersive virtual reality equipment [5,14]. In four of the studies, participants were playing exergames [2,6,9,10], meaning that as they moved, the avatar moved with them. In the remaining studies, participants controlled the avatars with either a console controller or keyboard and mouse.

Results Summary

There is no strong evidence for an effect of avatars on health-related behaviours. The most reliable finding was related to the Proteus Effect, and suggested that larger bodied avatars would result in reduced effort during exergame play.

However, the populations between these effects were heterogenous, consisting of men only, women only, and male and female 'overweight' children. Insufficient procedural information was available to form judgements on risk of bias.

4.7 Discussion

The purpose of this review was to assess the evidence for avatars influencing health-related attitudes, and behaviours. The original aim was to estimate meta-analytic effect sizes from existing studies however there were too few and the variation between studies was too great for this to be informative. Instead, the main findings and quality of the studies will be assessed qualitatively.

Evidence of Effectiveness

Assigned Avatars

The most consistently successful paradigm was the manipulation of avatar body size to affect in-game performance. However, this was only tested in three studies, with the remainder of 'obese avatar' studies testing post-game behaviours or virtual shopping decisions. Two of the studies used identical methods (virtual tennis game) on college students of different genders, and the third was conducted on a cohort of larger bodied children, and utilised a running game on the Nintendo Wii. To say that this effect is reliable would be premature, and substantial work is required to confirm this. The heterogeneity in the samples used is an area that requires some discussion, since it indicates different purposes of the study. In the two studies by Peña et al [8,9], people were selected based on being in the "normal" Body Mass Index (BMI) range. The intention appears to have been for the participants to perform downwards comparisons on the larger sized avatar, and lateral or upwards comparisons on the smaller avatar. In the study by Li et al [7], participants were selected based on being in the higher BMI range, and would therefore be expected to perform lateral comparisons on the larger avatars, and upwards comparisons on the smaller avatars. The absence of baseline performance in any of the studies means that it is unclear whether any of the participants were performing or self-reporting at, below, or above their *true*s scores.

In broader terms, the evidence for the effect of avatar physique on the various outcomes. Only one study found significant effects of avatar body type on attitudes, intentions and behaviours. In the remaining studies, these measures were either not significant [14], or not measured [3,5,9,10]. Moreover, exposure to normative ideals, and hyper-ideal avatars had small or mixed effects on body image related outcomes. For instance, muscularity had no effect on body satisfaction in males in study [8], but did on one scale in study [13]; and in study [6], women self-reported more favourable attitudes towards their weight, but less favourable body satisfaction. Taken as a whole, it is unclear whether merely controlling an avatar will affect psychosocial constructs, whereas the effect on behaviour during play may warrant further investigation.

Customised Avatars

Evidence of the effect of the customisation of avatars on health behaviours was also mixed, with the direction of effects varying both between and within studies. Three studies directly compared assignment with customisation. Only one study reported a significant difference between the two conditions: Kim and Sundar [4] demonstrated that participants who customised their avatars had higher self-reported behavioural intentions, as measured by the percentage of time they intended to spend on maintaining good health (when controlling for body discrepancy), but lower observed behavioural intentions, as measured by a coupon selection method. Correlations between the two measures were not reported, but if they were measuring the same construct, then the direction of the effect ought to be the same. Elsewhere, customisation has been shown to improve a user's identification with their avatar (Turkay & Kinzer, 2014), which in turn has been shown to influence motivation to play games (Birk, Atkins, Bowey, & Mandryk, 2016), enjoyment and exercise intentions during an exergame (Li & Lwin, 2016). However neither study measured identification, or related constructs such as self-presence, which would be beneficial in future studies.

The type of instructions given to participants customising their avatars was manipulated in four studies. The instructions were based on self-concepts and models described in Self Discrepancy Theory (SDT; Higgins, 1987), namely actual, ideal, or ought selves. Mere customisation based on these concepts had null [1,2], positive [4] and negative influences [11] on health-related outcomes. Since there is no discernible pattern, the distribution and conditions under which these results were yielded is of interest.

The actual self is the perception of the self as it currently is. In Study [4], Kim & Sundar hypothesised that the creation of the actual self would lead to higher perceived risk to the body. They found an effect, but three scales were used and only one supported their hypothesis with no theoretical reason provided for choosing one over the other two. Study [1] also used actual-self activation but found no significant differences between the actual and ideal avatars on any outcome (perceived risk, self-efficacy, behavioural intentions). Other studies used actual self-activation as a baseline [2,10].

The ideal self is a projection of who a person desires to be. In studies comparing actual and ideal selves [1,2,4,11], there were no significant differences reported between groups on any health related outcomes.

The ought self is the projection of who a person feels they should morally be. This self-model was only used in one study [11], and the authors reported that when controlling for health consciousness, participants selected fewer food items compared with those in the actual body-activation condition. What the authors do not discuss is that when not controlling for health consciousness, and when all participants were analysed together, the ought self had a stronger, and significant negative effect on the observed quantity of healthy foods eaten.

However, none of these studies actually measured self-discrepancy to see if the avatar exposure had actually worked in the intended way. The inconsistency and low quantity of results indicates that more work is needed in this area, as it is not clear whether customising avatars has any effect on health-related outcomes.

Quality of evidence

By clinical trials standards, the quality of the studies was inadequate, with insufficient information being provided to form valid judgements in the majority of bias domains. This is likely to be a problem with the field in general, which is starting to catch up with reporting standards, but ultimately it means that no conclusions can be made about how the studies were run. In three studies, participants were asked to report on the nature of the study, and those who guessed the purpose were excluded from further analysis. However, no information about the blinding of conditions or allocation concealment was provided, with only a cursory mention being made to the randomisation of participants. Two studies mentioned that allocation was performed prior to the participants arriving, which could indicate a higher risk of selection bias. No studies reported conducting an *a priori* power analysis to inform their sample size, nor was a smallest effect size of interest (SEOI) declared, with the majority of studies conflating significance with relevance. Power was only mentioned by Jin, who reported an *a posteriori* 'observed power' after the sample had been collected and the effect size estimate observed. It has been convincingly argued that observed power provides little information of use by meta-scientists and statisticians, because the observed effect may be inflated due to outliers having a disproportionate influence on the overall estimate (Baguley, 2004). Only one study included a within subjects design looking at pre and post measures for vaccine intentions, self-efficacy and risk perception (Darville et al., 2018). This is important because there is no indication in most studies of whether participants were already at a ceiling level for the outcomes. If they were, then there would be no way for the manipulation to improve outcomes in certain participants.

The validity of measures used within the studies was not discussed in any of the articles. This is important because few of the studies directly measured behaviours, and most relied on the estimation of psychosocial constructs. Without a discussion of the construct validation of these measures, it is impossible to judge whether the items are appropriate for the purpose (Cronbach & Meehl, 1955; Flake & Fried, 2019) Some studies used standardised measurements, but this does not seem to have had any influence on the outcomes.

One area in which risk of bias was high was attrition. The degrees of freedom reported by the majority of the studies (N = 14) do not match the reported sample size, suggesting that the analyses were run on different samples, with no discussion of attrition. Although these differences may be due to typos, Statcheck was run on all inferential statistics and no errors were detected (Epskamp & Nuijten, 2015).

Sources of heterogeneity

At the most general level, this study investigated the use of avatars to influence health-related outcomes. However, there was considerable variation in the uses of avatars to affect such outcomes. The level of interaction with the avatar varied between the studies in the set. In nine of the experiments, participants had no control over the appearance of the avatar, and so any self-similarity would be coincidental. In two of the studies, participants customised but did not control their avatars. In the remaining three studies, participants selected [1], customised [1,2,4,11], or were assigned [4] avatars. All customisation instructions involved some level of self-similarity, that is participants in these experiments were limited in their freedom to self-present.

The choice of dependent variable was also a large source of heterogeneity.

These ranged from self-report to direct observation, from post-treatment to peri-treatment behaviour. The domain of the dependent variable did not always match the treatment. For example, in one case [2], the treatment was an exercise game and the outcome was healthy eating intentions. An argument could be made for the higher success of studies at finding an effect when there was a close, theory-driven, correspondence between the dependent variable and the treatment. The Proteus Effect predicts that stereotype-relevant behaviours will be adopted when an avatar visually represents this group. If this were to follow into the muscularity studies, Proteus Effect would predict variations of in-game behaviours, rather than post-play attitudes towards the self.

A further source of heterogeneity was the reason for conducting the study, and the implicative chain that led to the hypotheses of the studies. The Proteus Effect formed the key avatar theory in 9 studies [2,3,4,6,9,10,11,12,13], whereas virtual self-modelling did in one study [12], and four did not cite any of the main avatar-influence theories [1,5,7,8]. In addition to these avatar influence frameworks, there was a large variety of ancillary theories attached to the chains of many of the studies. The most common was Self Discrepancy Theory [1,2,4,5,11], but Social Comparison Theory was mentioned in 5 [4,7,8,9,10], stereotype threat induction was used as a second factor in one study [6], as was Regulatory Fit [2], and Social Cognitive Theory was used by four studies [2,3,4,9]. The use of covariates was inconsistent throughout the dataset, again varying between and within studies, as well as not featuring in the reported *a priori* hypotheses of many studies.

Sources of Homogeneity

Discussing heterogeneity in the reviewed studies serves the function of identifying sources of variation that may explain differences in findings across studies, and potential future moderators. In discussing homogeneity, the elements of the set of studies that do not vary sufficiently will be discussed. This may be due to the over-reliance on convenience sampling in psychological research. In a rather scathing critique of psychology[5], Henrich et al. (2010) note that participants in 96% of Psychological studies are sampled from 12% of the world's population, and note variability between WEIRD (Western, Educated, Industrialised, Rich, Democratic) countries and non-WEIRD countries in a large range of common psychological traits. For example, such differences extend to self-concept, with WEIRD denizens being more likely to describe themselves in terms of the self (traits, personality, etc), and denizens of some non-WEIRD countries describing themselves in terms of relations with other within their society (Henrich et al., 2010), with convergent evidence for this offered by Morling & Lamoreaux (2008), who show that cultural products such as media portrayals and advertising are more interdependent in non-WEIRD countries than WEIRD countries.

Differences in the degree of self-enhancement have also been noted, with the drive to see the self positively being a distinctly western urge (Heine & Hamamura, 2007); it is feasible that this would result in cultural differences in avatar customisation, particularly in terms of ideal and ought selves. Other differences include motivation to conform (WEIRD > !WEIRD) which may affect responses to interventions based on the ought self (i.e. pressure to conform to the expectations of others); and variations in Psychological essentialism (the attribution of psychological identity to other living beings), which could feasibly influence avatar perception (for instance, the degree to which an avatar is seen as *an other*, see page 22). Any of these factors may affect the nature and extent to which avatar appearance affects the player. This is before considering cultural differences on avatar use or relations, the patterns of which vary not only between WEIRD and Non-WEIRD cultures, but within WEIRD cultures too (Banks et al., 2019).

The studies within the current review were predominantly conducted in the USA (N = 9) with the remainder distributed between South Korea, Japan, Singapore, and Taiwan. Additionally, in all but one of the studies discussed in this

[5] Although the critique could easily be applied to any number of fields.

review participants were drawn from undergraduate student populations, and the maximum age reported in a study was 34 years. The remaining study investigated children between the ages of 9 and 12 years [6]. That is, there were no studies conducted on adolescents, middle aged and older populations, despite the fact that these represent non-trivial proportions of the gamer demographic. The overuse of undergraduate students is a pervasive problem in psychology, and is usually addressed in a perfunctory sentence in the limitations sections of papers. However, such reliance has implications for the external validity of findings. To paraphrase Blastland (2019), what seems like knowledge in one context may be ignorance in another. Undergraduate students have many characteristics that may affect their performance. First of all, they are usually students from the department in which the study is being conducted - e.g. psychology students. By virtue of this characteristic, undergraduate participants are unlikely to be naive to the purposes of the study, even if they are not familiar with the particular paradigm. Next, USA undergraduate students are unlikely to be representative of the general sub-population let alone the species (Arnett, 2016); they are younger, come from affluent backgrounds, and are more likely to be white ("Young adult educational and employment outcomes by family socioeconomic status," n.d.). That is, even though there was heterogeneity within the studies in this review which threatens generalising about the internal consistency of any effect of avatar-appearance on health behaviour, the homogeneity of the samples threatens the generalisation to humanity as a whole. A good solution to this issue would be running more large scale, theory driven research in a wider number of cultures, such as attempts made by the Psychological Science Accelerator project (Moshontz et al., 2018).

Strengths and Limitations

Although there have been at least two qualitative reviews on avatars and health behaviour (Ahn, 2015b; Ahn & Fox, 2017a), this study is the first to systematically collect and assess the evidence of avatar appearance on health-related attitudes, beliefs, intentions, and behaviours. These reviews presented preliminary findings on how doppelgängers can influence health behaviours, as well as how these may be integrated with social psychology theories to produce favourable health-related outcomes. The current review has assessed studies investigating the use of avatar appearance to influence health behaviours and addressed methodological and theoretical flaws. It is hoped that it will be used in conjunction with the existing reviews for future avatar researchers to plan their studies.

A strength of the study was the use of an established risk of bias tool. However, the Cochrane Risk of Bias tool is designed for clinical trials, and much of the information that would be provided to support low bias (e.g. randomisation protocols and measures to reduce allocation concealment) is not commonly included in psychology and human factors research.

One important finding was that there are surprisingly few studies on this topic, which meant that the estimation of a meta-analytic effect size would not have been informative.

The protocol for the review was registered, but the results vary from the stated plan. For example, the plan was heavily focussed on the use of meta-analytic techniques, with little space dedicated to contingencies should there be insufficient data available. Further, the outcomes that were pre-registered were intentions, attitudes, body satisfaction, and post-intervention self-report and observed behaviours. During the selection phase this was adapted to include in-game behaviour, since this was a common outcome in studies captured by the search terms. It is also important to note that, although various authors were contacted regarding file-drawer studies, and non-peer reviewed sources were searched (e.g. ProQuest thesis repository), there were no unpublished studies found, although one study had previously been a master's thesis. This is a limitation because given the near ubiquity of publication bias in Psychology (Simonsohn, Nelson, & Simmons, 2014), and in videogame violence research (Ferguson, 2007), it is likely that this review does not contain all conducted studies in this field. Further, the inclusion criteria for the studies had to be adapted, since the pre-registered criteria were at once too broad and too narrow. That is, they stated that any study using an avatar would be included, and that outcomes had to be measured before and after the intervention.

4.8 General Discussion

Findings involving impacts of avatars on health-related attitudes and behaviours were mixed, and there was not enough data to draw firm conclusions. Perhaps the most consistent result was the positive effect of playing exercise game as an obese, compared with "normal sized" avatars on in-game performance, with three studies reporting statistically significant results on this effect. The three studies were conducted on different groups, with one study demonstrating this with children, one with adult men, and the last with adult women. There have been no published attempts to replicate these findings.

Taken at face value, a weak claim could be made for an inverted U shaped relationship between avatar appearance and exercise behaviour. The use of overweight and over-muscular/idealised avatars may reduce exercise performance, or body satisfaction respectively, while an optimal level of performance or persuasive enhancement may occur in-between the two extremes. There is not nearly enough data to support this model, and virtually no studies on "normal" body types. The range of body types used within the appearance related set of studies is restricted to the thin/athletic and overweight/obese sides of the spectrum. The models used in studies [9] & [10] were lean and muscular, and likely did not reflect an 'average' body type. The avatars in study [6] appeared overly skinny. Indeed, the definition of "normal" amongst studies is ill-defined, and would need to be established prior to further work being conducted. Further, the avatar body-image studies were conducted in different countries with very different cultural norms (Japan, Singapore, Taiwan, USA) and so there is an additional source of heterogeneity involved. Only one of these studies looked at intentions to exercise, and found that these were reduced with an obese avatar. As mentioned, there were no adequate controls for these studies - for instance an overweight avatar, played by an overweight child could be seen as “normal”, while a “normal” sized avatar is seen as ideal. This was not explored within this particular study, so the direction of the effect is not clear.

Similar issues are present with the avatar customisation studies. The dominant paradigm in this set of studies was Self Discrepancy Theory (SDT) whereby the priming of certain self-concepts leads to negative affect and depressive symptoms. In these studies, participants were instructed to create avatars after being primed with different self-concepts (actual self, ideal self, ought self). These studies generally hypothesised that this level of priming would improve participant’s health behaviours because they would be able to visualise an ideal self and strive to meet this. Again, findings are vastly mixed. Study [2] reported no main effects of avatar customisation on healthy eating intentions, whereas study [4] demonstrated that people who created their actual self-reported a higher perceived risk to the body than those in the ideal self-avatar condition. Moreover, the act of customisation was associated with higher self-reported intentions to perform behaviours, but lower observed intentions to perform behaviours (as measured by coupon selection). Finally, study 12 demonstrated that creating an avatar that reflected an ought self-reduced the number of healthy food items selected by participants after engaging in a virtual intervention, with no significant effects reported for the actual or ideal avatars. It is important to ground future studies using SDT concepts in the actual theory. The theory predicts that drawing attention to discrepancies will increase negative affect, which ought to result in a less favourable gaming experience, reducing any behaviour changing effects that a game or intervention may have. Previous research has attempted to use these self-concepts as motivational tools, but there is little aggregated evidence that this is the case, and potential deleterious effects ought to be examined.

Other studies have suggested that the mere customisation of an avatar may be sufficient to illicit attitude or behaviour change. However, no studies have actually demonstrated a main effect of avatar customisation on health related measures, showing only statistically significant interactions, or decreased health outcomes.

Two studies [5,12] investigated the effect of the 'future self'. These used manipulated images as avatars, with study [5] reducing the weight of the avatar to the participants ideal, and study [12] adding smoking-related age artefacts to the face. These studies yielded fair results and there has been some compelling research on appearance related interventions for a range of health behaviours including UV protection (S. Persson, Benn, et al., 2018; S. Persson et al., 2018) and smoking (Grogan et al., 2011), as well as applications in non-avatar based virtual reality studies (Ahn, 2015a). As such, this may be a fruitful direction for future research.

Grounding in Theory

As with any line of research, in moving forward with avatar-related health interventions, it is important to both progress and to ensure that what is already known is incorporated into future research (Lakens, 2014).

No studies had measures for avatar identification, which has been shown to be important in influencing intrinsic motivation to play games (Birk et al., 2016), and can mediate exercise intentions in exergames (Li & Lwin, 2016). Interaction with avatars has been discussed in terms of connection and disconnection, this could be explored further (Ratan & Dawson, 2016).

Proteus Effect

In general terms, the mechanism behind behavioural assimilation from avatars is described by the Proteus Effect (Yee, 2007; Yee et al., 2009). Under this theory, embodying an avatar results in the user drawing inferences from the appearance, and behaving in a congruent way. It is not sufficient to simply state that changing an avatar's appearance will result in health behaviour change, since the Proteus Effect has a number of necessary conditions that must be met for behaviour to change. To wit, there must be a visually activated stereotype (positive or negative) associated with the target behaviour; the avatar must unambiguously represent this stereotype; the stereotype must be general enough that participants will have encountered it. Only three studies in the set met these conditions, and these were the three that found significant results on in-game performance. There is a wide literature on weight stigma and a persistent belief that larger individuals are lazy, slow, and lack control (Flint & Reale, 2018; Kowert, Festl, & Quandt, 2014; Seacat & Mickelson, 2009). It is perhaps unsurprising then that the manipulation of avatars to have larger, stigmatised body-types would affect behaviour, if the Proteus Effect holds. It is interesting to note that the health related outcomes did not consistently extend to real world behaviours. This is consistent with the original Proteus Effect studies which had weaker, un-replicated evidence for post-experience effects. There is no literature on stereotypes related to idealised and hyper-idealised body types, and the studies that attempted to manipulate body satisfaction using avatars relied on traditional media influence theories. Such studies have been shown to be weak, and only have an effect in very specific circumstances (e.g. women with pre-existing body image dissatisfaction) (Ferguson, 2013). It is therefore unlikely that muscular, or normative idealised avatars would affect people self-reported attitudes under this framework. That is not to say that hyper-muscular or idealised forms will not affect attitudes or behaviours, but future studies should meet the necessary conditions of the Proteus Effect, and draw theory-based hypotheses first.

Self-Discrepancy

Similar arguments to those described in section 4.8 may be pitched at the avatar customisation studies. In general, the studies aimed to promote health related behaviours using self-concepts from SDT. The chain of reasoning in, for example, Kim and Sundar (2012), is that when people create an ideal avatar, they observe their distorted self-perceptions from the third person perspective, which results in reflexive self-perceptions, carrying over into their behaviour as per the predictions of Self Perception Theory (Bem, 1972). This chain of reasoning is not clearly supported by the literature. Higgins (1987) states that Self Discrepancy Theory also has necessary conditions for activation of a discrepancy. A particular self-concept must exist, and be accessible. At no point in any of the customisation studies was self-discrepancy in participants assessed. Self-Discrepancy measurement typically has an ideographic and nomothetic component, that is the participant states their ideal, ought, or feared attributes and then indicates the distance their actual self is from these attributes (Hardin & Lakin, 2009). The simple inclusion of such a measure would allow researchers to predict the avatar influence behaviour more precisely. Participants were provided with instructions to create avatars that represented their actual, ideal, or ought selves, without determining whether such discrepancies existed, or what the nature of the discrepancy was. The assumption within the studies was that participants would make a healthy version of themselves, but their ideal could just as easily be someone who is great at videogames. Without determining whether a person has health behaviour internalised, it is not possible to know what the avatar actually represents to them. In Kim and Sundar's study, participants were simply asked to "customise their avatar to resemble their [actual/ideal] self" - and so then if there is an underlying avatar persuasion mechanism, it is unclear what is being activated when the avatar is used, and which behaviours mapped onto the representations.

Applied Context

This study has identified that there is insufficient quantity and quality of information currently to use in the applied context. The literature is either inconsistent or laden with value judgements. The findings of this study may be applied to the meta-context of the field: that is in its current state, the avatar appearance literature should not be applied to real world contexts. In section 4.8, suggestions are made to tighten the avatar-appearance persuasion literature and place it on a progressive research line, which will help make the field fit for the purpose of informing policy and development.

Findings and Conclusions

Since exergames often provide direct control of an avatar through movement of the players body, and might be a powerful source of visual feedback (e.g. Fox & Bailenson, 2009), it is perhaps surprising that more studies have not been conducted on the effect of avatar-appearance on in-game performance. Notably, no studies utilised the Wii Fit hardware and software which offers vicarious reinforcement through avatar weight changes. As briefly discussed in Chapter 2 (page 19), Wii Fit hardware has been associated with positive outcomes for enjoyment in *obese* samples, and with intentions to exercise in general (Garn et al., 2012), and attempting to extend these effects using avatar appearance may be an interesting avenue.

This review identified a range of ways that avatars have been used to affect health-related outcomes. Despite the assertions of Nowak & Fox (2018) that avatars may represent a user in any modality (page 20), the avatars used in the studies described were exclusively visual, predominantly three-dimensional, humanoid, and animated (the avatars used in the Darville study were two-dimensional static images, and the avatars in Sah et al, and Song et al's studies were both two-dimensional). The uses of avatars included customisation to reflect ideal, actual, and ought selves (discussed later, in Chapter 5, Section 5.2); time progression studies; prescription of larger-bodied or athletic avatars[6]. Most pertinent to the current thesis are the latter, which mostly follow the Proteus Effect paradigm (page 26), relying on stigma and stereotypes to fuel the mechanism. Only a handful of studies utilised the Proteus Effect theory, and only three supported this. Arguably, the three *successful* manipulations were closest to the necessary conditions discussed on page 52, with behaviours (physical exertion) matched to the salient stereotypes through the invocation of the $laziness|bodysize$ activation. The study by Verhulst et al. (2018) unsuccessfully attempted to invoke $poordietchoices|bodysize$ responses using immersive an 3D environment and a shopping task, which may in part explain why the Verhulst et al study was unsuccessful.. Using Sims characters, Joo & Kim (2017) attempted to invoke $highcalorieeating|bodysize$ and $laziness|bodysize$ responses, both of which occurred after the game, and neither of which were successful. The issue here is twofold: first, participants were not given the opportunity to perform behaviours whilst in control of the avatar; and the use of Sims characters suggests an observational involvement, rather than an embodied relationship with the character. In Chapter 2 Section 2.5, the arguments between Peña and Yee were highlighted and it was concluded that control, as well as observation of the avatar was an important component of the Proteus Effect (see pages 31 and 49). Ignoring other considerations such as false negatives, this may explain the variation in findings within the set of Proteus Effect studies.

Of the effects discussed in this review, the most consistent required the presence of either weight stigma, or athletic reverence in participants to meet the necessary conditions of the Proteus Effect. Participants in the Li et al (2014), Peña (2014,2016), and Verhelst (2018) studies were required to associate the larger-bodied avatar with *out-there* stereotypes (discussed on

[6]The research question was specifically about avatar appearance so this might seem tautological, however during the filtering process there were no non-visual avatar studies either.

page 56) associating this group with a lack of control, greed, and sloth. As discussed in Chapter 2 (from page 44), weight stigma is pervasive, occurring in many different cultures globally, and it also has specific ”predictions” of how a person with the attribute might behave, making it a good candidate for a Proteus Effect demonstration study. However, the authors go beyond using their findings for mere demonstration, and make suggestions for game design decision which, with such weak evidence, are premature. Arguably, promoting stigma, even if it is for such a *just* cause as promoting weight reduction, is counter-productive and potentially harmful. Regardless of the intentions behind these previous studies, there are many demonstrative studies supporting the Proteus Effect, which makes it the most promising foundation for a progressive research line of the theories discussed in this review. What was clear, however, was that the previous health-related Proteus Effect studies did not have sufficient bias-reducing measures in place, nor were the avatars used validated for the purposes of the study. This will inform the development of methods in the later chapters of this thesis.

Conclusion

In conclusion, there is insufficient evidence to support the use of avatar appearance as health-related interventions, and suggestions of limiting avatar body sizes to athletic or ”normal” ranges are not supported by the aggregated body of knowledge. Careful studies and replication attempts are required to explicate the interaction between avatars and real world health-related behaviour. That being said, three studies supported the use of athletic avatars to promote increased physical exertion. Moreover, a variety of review articles interpret the previous studies as strong evidence in favour of avatar appearance as an influential factor in weight management and health-related behaviour settings (Horne et al., 2019; Rheu et al., 2019), as well as a variety of other non-health related behaviours (Ratan et al., 2019). Given the inconsistencies in interpretation of prior findings, the following chapters will critique both the suggestions of Li and Peña's studies, and the Proteus Effect in general. This critique will involve a plurality of methods, including qualitative and quantitative techniques.

Chapter 5

A Sort of halfway house, a hero version of me

5.1 Rationale

Chapter 4 argued that, from the limited available studies, physical activity appears to be the health-related behaviour that is most sensitive to avatar appearance. Moreover, altering the athleticism of an avatar had the most reliable effect on the intensity of physical activity, relative to other possible interventions (e.g. ideal-self motivated avatar customisation). Some of these studies included rather draconian suggestions advocating the limitation of avatar design options, restricting avatars to thin body types. In line with the second objective on page 5 (explore the experience of playing an exercise game with a facially similar, athletic avatar), this chapter approached the research question on page 58 from a qualitative perspective. Participants were restricted to avatars with athletic body types and interviewed about their experiences. The study will also discuss claims made in the exergaming literature reviewed in Chapter 2 (page 17) that exergames might be a viable alternative to *traditional* exercise; discuss this in the context of the findings from Chapter 4 , and critique the claims made by authors of the studies therein.[1].

5.2 Introduction

In Chapter 4, a systematic review of the literature presented the findings of studies in which participants using plus-sized avatars moved more slowly or made fewer movements than those using a "normal" sized avatar (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014). This led the authors of these studies to either indirectly (through "optimising virtual environments for exercise"), or directly suggest that the number of body-related affordances[2] available to players ought to be limited to the 'normal' Body Mass Index (BMI) range, as is shown in the following quote from Li et al. (2014):

> As exercise attitudes and performance have been shown to decrease when overweight children are assigned overweight body weight avatars, giving these children a fixed avatar that is bigger in body size may not be a good idea if the purpose is to encourage them to exercise. [p45]

Although it was released before Li's study, the sports exergame *Kinect Sports Rivals* (Rare (2014), KSR) does just this. Players of KSR create an avatar using a system that restricts the range of available body types to *heroic champions*. In this chapter, the experiences of gamers playing with avatars that were automatically generated to look similar to them, but with appearance-affordances limited to athletic body types, were explored using qualitative methodology.

[1] The pre-print for this article may be found at https://psyarxiv.com/6ensp/.
[2] parameters that a player can customise.

Avatars and Affordances

Affordances are the number and range of parameters available to a player when customising the appearance, traits, or personality of an avatar (McArthur, 2018; Nowak & Fox, 2018). These may define things such as skin tone, eye and hair colour, facial features, body shape and size, intelligence, and charisma. The affordances of a game define the representation space accessible to the player and is susceptible to the cultural and ideological rhetoric of the developers (Kolko, 1999).

For instance, the Nintendo Wii *Mii* creator has been criticised for being influenced by visual culture, having fewer affordances for the representation of black people (Higgin, 2012; McArthur, 2018). Similarly, Kafai et al. (2007a) report frustrations of their teenage participants at the limited range of options for creating non-white avatars in the online community WhyVille.

If suggestions from Human-Computer Interaction authors, such as those discussed in Chapter 4 are followed, then there is a danger that similar frustrations will be experienced by those who do not have bodies that conform to the normative slim and muscular ideals. Videogames are already relatively limited in their representation of bodies. A series of content analyses identified that the average male and female videogame characters represented normative ideals, and that this was becoming more apparent as the capabilities of videogame systems to depict photorealistic characters increased (Martins et al., 2009, 2011). The positive representation of larger bodied people in both print and digital media is also extremely limited, and these individuals are often depicted in anonymous or stigmatised ways (Heuer, McClure, & Puhl, 2011; Pearl et al., 2015a, 2015a; Puhl et al., 2013)

In addition to promoting weight stigma by reducing the availability of different bodies in games, there is some research that suggests exercising with ‘real life’ thin or athletic ‘fit’ others may have a deleterious effect on exercise intentions (Wasilenko et al., 2007); and exposure to idealised media depictions of people may increase body dissatisfaction in at risk individuals (Ferguson, 2013; Robinson et al., 2017; Tiggemann & Zaccardo, 2015). This has mostly been explored quantitatively, although in a qualitative presentation of participant's experiences of gym attendance by Pridgeon & Grogan (2012), there was some discussion of positive and negative upward physique comparison with athletic others. To the best of The Author's knowledge, this study will be the first to explore such physique comparisons in a virtual environment.

Avatar Customisation

To date, there has been a large body of work on the purposes that customisation of avatars may serve (Mancini & Sibilla, 2017). For example, avatars have been shown to be used as an externalisation of the self (Turkle, 1997); a doll to dress up (Liao, 2011); or a product or consumable object (Cui, Aghajan, Lacroix, Halteren, & Aghajan, 2009).

Two approaches to avatar customisation have been posited: relational and socio-constructivist (Castronova, 2003; Mancini & Sibilla, 2017). Under the relational approach to avatar customisation, users aim for some semblance of similarity with the self when designing their character. However, the flexibility in avatar customisation means that users may enhance the representation such that it aligns with their ideal self. The disparity between the actual and the virtual self is called Virtual Self Discrepancy (Jin, 2012) who found that this construct was with reduced self-presence (discussed in Chapter 2, page 25), and flow (Nakamura & Csikszentmihalyi, 2014). Under a social constructivist approach to avatar design, users take the opportunity to explore new identities and forms (Vicdan & Ulusoy, 2008). This may include alien beings, a wider range of genders, or malevolent anti-heroes (Mancini, Imperato, & Sibilla, 2019).

Similarity to, and indeed enhancements of the self, may be achieved through psychological or physical attributes (Messinger et al., 2008). There is only limited qualitative research on virtual self-enhancements (Messinger et al., 2008), and few attempts to explore these enhancements and how people experience their own bodies (Cacioli & Mussap, 2014). These qualitative approaches were primarily used to supplement quantitative findings, and responses were gathered using open ended survey questions. The current study aimed to get a richer representation of participant experiences though in-depth interviews.

Naturally, what counts as *enhanced* will differ between people and will depend heavily on subjective norms. The current study will critically address the recommendations made by Li et al (2014) and Peña et al (2014, 2016) to limit affordance parameters to “normal” BMI and athletic body types, by discussing the experience of playing an exercise game with an athletic representation of the self. Discourses surrounding this experience will be contrasted with the reported decisions made by players when designing avatars in their own gaming lives.

Exergames and Motivation

There is evidence that exergames can be motivational for exercise in the quantitative literature (Li & Lwin, 2016; Limperos & Schmierbach, 2016; Van Nguyen et al., 2016). Li & Lwin (2016) propose a so called ‘Avatar Motivation Model’ of exercise, in which players who experience the illusion of self-presence (i.e. feel as if they are ‘in’ the game) identify more with their avatar, leading greater to enjoyment which increases motivations to play the exergame, which in turn mediates intentions to exercise. Qualitative inquiry into the motivational aspects of a commercial exergame will allow a richer and deeper exploration into the features of exercise games that players feel they would most enjoy.

Aims and Objectives

In the current study, a qualitative approach was taken to identify the goals that participants have when creating avatars, and how they responded to being represented by an avatar which is constrained to having an athletic body type.

Broadly, the objectives of the study were to explore

1. What goals and strategies do people have when designing an avatar to look like themselves?
2. What are peoples impressions of an avatar that is generated to be an idealised version of themselves?
3. Is body image reflected in participant's accounts surrounding motivations and approaches towards avatar customisation, and reactions to the automatic representation?
4. To what extent do people feel the exergame was a persuasive experience, and may be motivational for health-related behaviours?

5.3 Methods

Design

The current study was an interview study during which participants created an avatar and used this avatar to play climbing and tennis sports games on an Xbox One exergame. The study was executed in four phases:

i. Participants were interviewed about their gaming behaviour, avatar use, health and body image attitudes.
ii. They then had an avatar automatically generated and were invited to interact with it in a virtual mirror (i.e. the avatar was displayed on screen facing the participant and reflected their movements).
iii. Participants then played a climbing and a tennis game with the researcher.

iv. Finally they were interviewed about the experience and how it may have affected their body image and health attitude.

During the avatar creation and gaming sections of the project, the audio recorder was active so participants could voice their impressions of the experience as it occurred. The interview sections of the session were semi-structured with pre-determined questions included on prompt cards (see appendix https://osf.io/he6us/) but with an opportunity for either the researcher or participant to expand on issues they felt were relevant.

Participants

The sample of participants consisted of 20 individuals (12 identified as female), with an age range of 20 to 42. These were people who fell within the gaming population of interest (gamers between 16-44, see from page 14). They were recruited from the Department of Psychology research participation pool (Sona), a public research participation pool (Call for Participants), and emails sent to staff and students by members of the research team. People who play or have played videogames that use avatar customisation were targeted for their experience of creating and using avatars. This was because they needed sufficient experience to speak knowledgeably and at length about their experiences. Sample size was based on the upper end of recommendations of the authors of the method (Braun & Clarke, 2013) and the need for themes to be generated across multiple cases.

Despite an effort to recruit a diverse sample by advertising through multiple channels, there was an over-representation of the Western, Educated, Industrialised and Democratic (WEIRD) population. However, given the small size of the study this homogeneity can lead to greater focus and stronger identification of meaningful themes (Braun & Clarke, 2013). The aim was to keep an approximately even Male/Female gender split since experiences of body image satisfaction are reportedly different (Grogan, 2017). The pseudonyms and available demographic data for each participant can be found in Table 5.1.

Inclusion criteria

The key inclusion criteria for the study was that participants had some experience of playing videogames. With an estimated 33 million gamers in the UK (IAB, 2011) the range of motivations and uses of videogames will be varied and the aim was not to only focus on 'hardcore' gamers or people who identified with the label 'gamer' but cover a range of levels of interest and interaction with videogames. Although the study is about exergames, experience with playing these games was not an inclusion criterion since the expertise of the participants in what they find to be a fun game was more important. That is, it was their critical perspectives as gamers that were important rather than domain specific exergame knowledge.

Apparatus

Avatar-based Videogame

Avatars were generated using an Xbox One videogame console with a Kinect camera using the Kinect Sports Rival videogame (Rare, 2014) on a wall mounted 43-inch HD LED television. Kinect Sports Rivals is a selection of sports mini games including wave racing, target shooting, climbing, tennis and bowling. Participants used the Kinect camera to control an on-screen avatar using their body as the controller. The tennis and wall climbing mini-games were selected from the available mini-games because they were the most active of the selection which included pistol shooting, jet-skiing, football and bowling. The avatar generation procedure involved recording a participant's body outline and facial features. These inform the creation of a 'Champion' avatar which approximates the users physical features. Tutorial videos for each of the games were presented to each participant.

Topic List

The choice of questions was informed by a Critical Realist epistemology, since the research question was formed under the assumption that there is a common, observable reality but the interpretation of this is necessarily influenced by the researchers experience (Willig (2013), p16). The questions were open ended and chosen to fit a 'funnel' sequencing method over three categories: videogames, health, body image. A funnel sequencing method involves starting each set of questions broadly (e.g. 'Tell me about your past and current videogame use') and narrowing down to specific elements (e.g. 'What specifically do you like about videogame avatar customisation methods?') (Tengler & Jablin, 1983). The after-play questions focussed on participant's experience of playing the game and how they believed a similar experience could affect the health-related behaviours and body image of themselves and others. These questions were also funnelled. During avatar creation and gameplay, participants were asked to think aloud in response to the process and experience. Occasional probes were made by the researcher, e.g. 'what do you think of the avatar', 'do you think it looks similar to you?'. Questions were designed by The Author and discussed with the supervisory team.

Data Collection

Data were collected over a period of 4 months between August and November 2017. Sampling adequacy was determined by the balance between appropriateness and adequacy of participants narratives (O'reilly & Parker, 2013). A minimum of 20 participants were interviewed to maximise depth of information obtained, and the decision to stop after 20 was made based on the content of the narratives from both interviewing and transcription. It was decided that the dataset contained sufficient information to answer the research question, but that is not to say that saturation had occurred, since the range of topics and the variety of experiences implied that novel information could have been produced in abundance (Malterud, Siersma, & Guassora, 2016).

Procedure

The study received ethical clearance from the faculty ethics board prior to data collection. All interviews took place in a 'simulated apartment' at Manchester Metropolitan University. This was an observational laboratory that had been designed to look like a living space, including a kitchen-living area, an inactive bathroom, and bedroom.

Participants gave their informed consent prior to the study beginning. The Dictaphone was set to record, and open-ended questions were asked until the items on the topic list had been covered, and participants declared that they had nothing more to add. Participants then had an avatar generated using the Xbox One. The process happened with the television muted to aid audio recording. The interviewer (The Author) led participants through the process in the absence of instructions. Participants were first positioned at a distance of approximately 2 meters from the screen while the Kinect camera device recorded some outline measurements. These were broadly based on height and body shape. Participants next indicated their age group (adult, N = 20; child N= 0) and gender (female, N = 11; male, N = 9). Participants were not restricted to same-sex avatars and were told that they could chose the gender that best represented themselves. Next participants were sat at a distance of approximately one meter in front of the Kinect camera while the software extracted their facial details. The avatar was then generated and revealed, participants were asked about their first impressions. In cases where the creation process had made obvious errors (e.g. as hair or eye colour), participants were given the opportunity to make changes. One participant [Miranda] was unhappy with the outcome because she felt it did not resemble her. It was reasoned that this may have been due to insufficient lighting, so the process was repeated with more lighting in the room. The same amount of lighting was used in all subsequent interviews. Once the avatar process had been completed, participants were given the opportunity to use the *Avateer* function. This was like a virtual mirror, in which their avatar reacted to their body and facial movements.

Participants were then shown tutorial videos included in the game in which a virtual coach gives instructions on how to play the mini games. Before play, participants were asked if they had any questions about the functionality. In both games, the interviewer played using their own generated avatar. A single game of climbing was played and up to five games of tennis. Participants were asked the post-play questions. During the post-play questions, a static image of their avatar was left on screen in the majority of cases (N = 18). Participants were then debriefed and were given an online demographics questionnaire to complete after the study.

Transcription

All verbal data were transcribed verbatim by The Author from the audio recordings at half speed, using ExpressScribe. The transcriber made notes and comments throughout the process, and the transcriptions were completed soon after the interview had taken place so that the context of the content could be recalled.

Data Analysis

The chosen methodological framework was Thematic Analysis, since the flexibility of the methodology was best suited to the ultimate goal of the study which was to model participant's experiences from the ground up rather than requiring a formal or prescribed philosophical or theoretical framework (Braun & Clarke, 2006). The method also allows for the interpretation of the data from a critical realist perspective, meaning that the presence and experience of the researcher in the process may be accounted for in a way that is not as feasible under more quantified qualitative methods. Moreover, this work was exploratory and as such as specific theory was not sought, nor was the purpose of the study to develop a model of influences on processes (Braun & Clarke, 2013a).

Analytic Process

Data were predominantly analysed at the latent level, with The Author interpreting participants' accounts in the context of his own knowledge about the field, activity, and literature surrounding videogames and avatar usage. The melding of theory and experience was inescapable, since The Author was neither naïve to the purpose of the study, nor the theories surrounding the research question. Some consideration of the surface content was allowed in the analysis - that is, if participants made relevant declarative statements, these were interpreted as facts, as well as contributions to latent manifestations. For instance, if a participant said "I exercise every day", that was treated as a fact, and used to interpret their other accounts. The Author intended to produce an inductive analysis, building inferences from the data and interpretations thereof, however some deductive elements were inevitable, given that the interview questions and research questions were developed *a priori*. In the rare cases where The Author reflected he was leading the participants using his own understanding, this content was interpreted critically and reflectively[3].

Completed transcripts (N = 20) were read twice before being imported into Nvivo 10 for Mac (Castleberry, 2014). The data were coded by The Author following the six-step process recommended by Braun and Clarke (2006) (Figure 5.1).

These were mostly surface content in the first iteration, with deeper patterns being coded on subsequent iterations. The coding was discussed with the supervisory team over 12 face-to-face meetings, by e-mail, and through comments during the drafting process.

[3]The main instance of this was at the end of the interview with Isaac, in which The Author defined technical terms from the cyberpsychology field, and the Isaac responded using these terms. This content was not included in the themes below

Phase	Description of the process
1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Figure 5.1: Screen clip of the 6 phases of Thematic Analysis from Braun and Clarke (2006), page 87.

Themes were then defined and named. The write up of the analysis was then checked against the data coding before the report was finalised.

Table 5.1: Pseudonyms and demographics of participants.

Pseudonym	Gender	Sexuality	Age	Ethnicity	Exercise	Gaming.Experience	BID	AI
Aaron	Male	Heterosexual	34	White British	Structured	Casual - Social		X
Alastair	Male	Homosexual	NA	NA	Precontemplation	Life-long	X	
Anwen	Female	Lesbian	24	White British	Structured	Life-long	X	X
Balthazar	Male	Heterosexual	27	White British	Structured	Former		
Cait	Female	Heterosexual	20	White British	Activity	Life-long		X
Cortez	Male	NA	22	White British	Activity	Life-long		X
Elone	Female	Heterosexual	23	White British	Structured	Casual		X
Franklin	Male	Heterosexual	NA	Mediterranean - Greek	Structured	Casual		X
Isaac	Male	NA	34	White British	Activity	Life-long		X
Keira	Female	Heterosexual	25	White Polish	Activity	Life-long		
Lily	Female	Heterosexual but Bicurious	NA	White British	Activity	Casual		
Meryl	Female	Heterosexual	22	White British	Activity	Life-long		X
Miranda	Female	Heterosexual	25	White Sweedish	Precontemplation	Not a gamer	X	
Nami	Female	NA	NA	White German	Structured	Life-long	X	
Nastasha	Female	Bisexual	28	White British	Structured	Casual		X
Nes	Male	Heterosexual	41	White British	Precontemplation	Life-long		
Nick	NA	Heterosexual	20	White British	Precontemplation	Life-long		X
Selphie	Female	Heterosexual	20	White British	Structured	Casual	X	X
Squall	Male	Heterosexual	NA	Chinese	Structured	Life-long		
Ward	Male	NA	NA	NA	Precontemplation	Former		

Table 5.1 continued

Pseudonym	Gender	Sexuality	Age	Ethnicity	Exercise	Gaming.Experience	BID	AI

Note. Categorisation was based on qualitative self-report// Structured exercise means that participants had a schedule and regular activities went to the gym and played sports. Participants in the Activity category exercised but would only participate in activities they found enjoyable such as wall climbing and pole fitness. Participants in the Pre-contemplation did not exercise in any structured way but claimed to be planning to do so. Participants were also categorised for gaming experience based on how they discussed their gaming history. Life-long gamers claimed to have always played videogames, casual gamers play socially but not habitually, former meant that gaming did not form part of their identity, but they had played game previously. BID indicates that participants expressed dissatisfaction with their bodies. AI indicates that participants identified with the avatar.

5.4 Results and Discussion

Presented are four themes. In brief, the themes describe the choices players make when designing their avatars in real life, and the functions that these representations serve and contrast these with the automatically generated idealised avatar. When the freedom of choice was removed and an idealised avatar was provided, reactions of the participants differed. This seemed to be informed by their body image, in that those who had internalised athletic ideals were more positive in their discussion of representation than those who had not. In Theme 3, the overall experience of playing the game and how this may affect their own, and other peoples motivations to exercise was discussed. Again, this seemed to differ based on existing exercise behaviour, with those who already exercised being less positive about gaming as a substitute for their own exercise, than those who were planning to start exercising. Finally, in Theme 4 I argue that there was an underlying 'fat-phobia' inherent in the accounts, with avatars being used to avoid 'fatness', and larger bodies representing player's 'feared selves'. However, this may be due to negative portrayals of larger bodied individuals in the media and may improve with wider representation.

Theme 1 - Body Image and Avatar Creation

During the interviews we discussed appearance-related ideals. This was phrased as participant's *ideal* body types. In a number of cases, these ideals were informed by societal norms, as participants discussed wanting thinner (predominantly, but not exclusively women) or more muscular (predominantly, but not exclusively men) bodies. Some participants [Cait, Cortez, Squall, Nes] described weight loss for health reasons as a factor in their ideal body. This was framed as 'not wanting to die young', or through rarely needing to visit the doctor. There was some evidence for the internalisation of 'normative ideals as morals' in the accounts, as words like 'should' and 'ought to' were used in reference to weight loss [Ness, Cait].

Body satisfaction in women (predominantly, but not exclusively), appeared to be low, with participants fluently discussing parts of their bodies that they were unsatisfied with. Women appeared to have a deeper internal dialogue with their bodies, compared with the men. The stories that the women described included the experience of previous weight stigma [Nami], experience of eating disorders [Selphie, Miranda], and body checking/ awareness [Nastasha] reflect the well-known societal pressures placed on women (Grogan, 2017).

There was a drive for muscularity in many of the men, with participants describing 'clear cut' or 'toned' bodies, however this did not manifest in dissatisfaction-related talk to the same extent as the women. Indeed, functionality of bodies appeared to be more important to the men in many cases [Balthazar, Aaron, Cortez, Franklin]. One male participant [Alistair] reported dissatisfaction with various parts of his body (e.g. legs, stomach), and appeared to have a negative experience of living in his body. In many cases, there was a discrepancy between how the participants wanted to look, and how they appeared to themselves. This was more prominent in women, but men also discussed their ideals in differential terms (e.g. wanting more muscle). It is important to note that participants were asked directly about their ideals, and changes they would make to their body, rather than this occurring spontaneously and so these responses may have been led by the researcher.

Discrepancy Reduction Avatar customisation was mentioned in relation to the participants' bodies for a variety of reasons, and most commonly to reduce the discrepancy between actual, and subjective ideal selves. The majority of participants said that their avatars reflected an ideal [Anwen, Alistair, Cait, Nastasha, Cortez, Nami, Squall, Elone, Selphie, Lily, Meryl]. Others said that they would generally give their avatars their actual shape [Franklin, Aaron], and one participant was not interested in customisation at all, choosing to use the default options instead [Isaac]. Avatar customisation was used by some participants to reduce the discrepancy between their actual and ideal bodies. Those who attempted to maintain their appearance but make minor ‘improvements’ were either more satisfied with their bodies or were already attempting to reduce the discrepancy through intentional health behaviours [Nastasha, Cortez, Anwen]. In these cases, avatars could be seen as a ‘quick fix’ for reducing self-discrepancy between the actual and ideal body types in those who are actively attempting to reduce this discrepancy. In line with the relational approach to avatar design (i.e. aimed for self-similarity when designing the avatar), these minor adjustments preserved the identity whilst providing the players with the body they felt they deserved, given the effort they put into changing their actual body.

> ANWEN I would always make it slimmer than I actually am, I, I don’t know what that is, um but I would always just keep it slim when in actual fact it should probably, well if I’m being brutally with myself I’d probably make it a bit bigger, but I think that’s more an aspiration of how I would like to look rather than an actual reflection of who I am.

Those who made these minor adjustments to their avatar’s appearance discussed this in terms of ‘vanity’. One of the motivations reported for creating a self-similar avatar was to put the player more in the game and experience things that otherwise would not be likely. For example, Franklin discussed using his avatar to experience playing in his favourite football team. For others [Nes, Alistair, Squall, Elone], part of the novel experience was adopting an entirely different body.

Some participants had ideals that were further from their actual body [Alistair, Nami, Elone, Selphie]. Their approach to avatar design aligned more with a socio-constructivist approach, forming new bodies through the avatar creation process, but was still anchored to their identity by way of ideal self-representation. These people avoided making themselves in game, opting for ‘attractive’ others, or people they would ‘like to be’. Nami discussed this:

> NAMI: my Xbox profile will look like me but the one in guild wars I realised was more like the ideal image you have of something, so it will be really tiny and really like assassin like, because I'm quite tall, and that was especially when you're younger, not a thing to aspire to as a girl, so like they were all like tiny and really thin and really athletic builds.

Her account also refers to pressures related to her height that she felt when she was younger. She went on to describe this process as "wishful thinking" which was a recurring comment in both those who designed their avatar to reflect their ideal selves, and ideal others. Alistair also discussed 'unchangeable' elements of his body (e.g. height and hair) as part of his ideal and reflected these in the creation of 'ideal other' avatars.

It was unclear whether it is limited digital affordances, or subjective affordances that lead players to present idealised versions of themselves. It is entirely possible that the two are so heavily correlated that this question is irrelevant. Social pressures were amongst the reasons for only creating 'thin' and 'attractive' avatars ["because society says so", Lily], as well as attractive avatars being more 'playable'. It is therefore debatable whether improving affordance availability would affect participant's projections, since they may also be subjectively confined to a limited, socially mandated space of 'acceptable' bodies.

Broadly, this discrepancy reduction approach is supported by the quantitative literature. Mancini & Sibilla (2017) argue that one of the goals of avatar customisation is to reduce the discrepancy between actual and ideal selves. Actual vs idealised selves, although this research was based on ideal and actual personality traits, rather than physical features. They also suggest that this is based upon player's offline personality, and that the type of avatar that a person will create will be informed by the type of player they are.

Other and Outré Avatars Indeed, the limits of affordances were generally explored by players who had a specific interest in creating grotesque or impossible figures. I refer to these as outré Avatars, and participants discussed these in fantastical terms, such as characters with unicorn tails and lizards with purple feathers. The motivations behind this included aesthetics ("I really like those purple feathers", Keira) or the approach to the game ("…or will it be a silly game and I'll just make people with purple hair and unicorn tails", Cait).

This approach highlights the 'playfulness' of avatar design and use, as discussed by (Vicdan & Ulusoy, 2008). Providing players with a diverse palate to draw from means that they can be more creative when designing their avatars. Outré avatars seemed to be reserved for exploration rather than canonical gaming. That is, participants played seriously or for fun. Serious gaming is planned, intentional, and would take up a non-trivial amount of time. For instance, Miranda discussed serious play in the Sims in terms of family generations, and Cait made the distinction between silly games and serious games. Both called for different avatars. For Nick, the avatar generation seemed to be a game in itself, but when playing for fun, he used non-outré avatars, some of which he made to resemble himself.

> NICK: mmm sensationalised, you go off the scale, try to make them look as ridiculous as you can like mad clothes that don't look right or just completely warped facial features, um I guess just mainly because it's entertaining isn't it to see how far you can stretch the limits of the creation system as it were, errrm, so there's not really any realism in it, it's all, yeah.

The minutiae of avatar customisation became less relevant the further from the actual self the avatar was. When creating themselves, people discussed the importance of tiny details, such like getting the "nose right" [Keira] when creating themselves, but not when creating their lizards. In these cases, increasing the sensitivity of the affordances, rather than the range may be helpful, although participants also discussed time constraints, and getting bored with the process ("I'd get to a point where I'm like ah fuck it , like you know that's a good enough nose", Miranda).

What was clear from the interviews was that some people do not have an interest in presenting themselves in games. The flexibility to explore different forms, or the provision of pre-made characters would be important for these people. This would either make the exploration of identities enjoyable, or to reduce the friction between switching the console on and being able to "get playing" [Isaac]. The crux of this point is that restricting affordances has the twofold effect of reinforcing 'moral' body types and removing the fun for those who like to create with the parameters. However, on the flip side of this point, some would have preferred fewer options. For instance, Isaac was not too concerned with the number of affordances in general. Moreover, for other participants, upon multiple playthroughs of a game, the avatar becomes less important, with emphasis being placed on exploration of the game world and abilities of the different classes of characters. A common model is to have a default figure, an avatar randomiser, and a wide range of parameters that can be customised (e.g. Fallout 4; Skyrim).

Theme 2: Choosing to choose your own ideal

Part of the interview was dedicated to discussing the idealised avatar that was generated by the Kinect Sports Rivals game, and how (dis)similar the avatar was to the participant. Despite the large number of participants who discussed the design of their avatars being informed by subjective ideals [Nami, Elone, Selphie, Nastasha, Cortez, Nick, Squall, Cait, Alistair, Meryl, Miranda], being presented as an athletic and muscular avatar was not a positive experience for everybody. Some described the avatar as being 'sexist' in its portrayal of the female body as a sexual object, with large breasts (Miranda, Nastasha); others noted that the avatar was idealised (Cortez, Nick) or over-idealised (Alistair, Nes), with one saying it looked like a 'jock' with a name like 'Chad, or Blake':

> NICK: …that smug, that smirk, is sort of like I feel like if I walked around looking like that someone would slap me.

Others saw it as relatively accurate (Nami, Balthazar, Aaron, Keira), or as an 'improved' version of themselves (Selphie, Anwen).

> SELPHIE: Err I feel like, yeah, it's quite. . . maybe accurate, I don't know actually I feel like it's a lot portrayed me a lot better than I actually am.

The lack of available body types was noted by some of the participants (Nes, Cortez, Alistair, Ward); and the fact that the avatar reflected some kind of moralised ideal ("what a male body should be", Alistair) was present in some of the accounts. This critique of the portrayal of people in videogames was informed in part by an appreciation of the potential harms that idealisation of bodies may inflict on people. Ward discussed how idealised portrayals of bodies could be harmful, particularly to women, and Alistair noted that they reinforced unrealistic standards. When self-representation was the goal, Ness said that there were two acceptable approaches: cartoony caricatures, or realism. He preferred more options with the body and suggested that the range of physiques was restrictive. He did not like the "halfway house" between realism and caricatured 'buffness' that KSR offered.

The avatar was not always seen as a subjective ideal; it was seen in some cases as an instance of a social norm, a prescribed ideal from some higher 'power'. This is can be related to work on the moralisation of bodies (Townend, 2009); word usage like 'ought' and 'should' suggest a moral standpoint; whereas 'like' and 'ideal' are more related to subjective norms. However, recalling that Lily created her pretty and attractive avatars 'because society says so', these subjective norms may be inherited from higher social norms.

Lack of Affordances The lack of affordances available to players was identified as a frustration, both in participant's own gaming, and in the portrayal of themselves in KSR.

Alistair noted that he was restricted to muscular forms in the games Skyrim and Dragon Age[4].

> ALISTAIR: Skyrim is a difficult one because you don't have any realistic body shapes, you go from thin and athletic to like buff hulk sort of thing and I think Dragon Age had even less control over the body.

The lack of variation in the avatar that was created for participants was noted as 'odd' by three participants. Ward noticed the homogeneity of body types between his own and the researchers avatar, and questioned whether this was something that was desired by 'people':

> WARD: I just noticed that they do look quite, like exactly the same body shape, both of them, don't they, they look quite muscular, yes, yeah, so did people want to see themselves as a muscular sporty person. Yeah, look at those arms!

I posit that this reflects the in-game rhetoric, in which the designers assume that this is what people want (Kolko, 1999; Nakamura, 2013). Indeed, this assumption was unwanted by several participants; it was deemed odd or strange at best, and sexist at worst:

> NESS: um, it's strange that they don't have any physique options, I'm not sure whether that happens automatically, whether it determines your body but it was they were all concentrating on the face and then you come out as this… umm highly buff, muscular version of yourself

[4]Both are fantasy role playing games.

with like big broad chest and you know an actual athlete but with your face.. errm which was strange, I thought they.. should have given the option to change your body type a bit.

MIRANDA: I think I'd like to say as well, her body is one of the reasons why I am quite sceptical of video games, like that is not my body, it's very kind of sexualised version of the female body. And it's almost like, why does the standard female, because that's the standard female body in a video game isn't it

Miranda also noted the gender differences between the avatars, saying that 'at least yours had muscles, where are my muscles?'. The women avatars were very slim and had small arms.

This is possibly another indication of the appearance related rhetoric within KSR. Men 'should' be muscular and buff, and women 'should' have a small frame, smaller muscles, and large breasts. Although she did not discuss this directly in reference to the generated avatar, Meryl enjoyed making avatars that had large muscles, an option that was unavailable to her in KSR unless she opted to create a male avatar.

The discrepancy between player and avatar was also evident in Squall, who did not subscribe to the mesomorphic ideal:

SQUALL: Mmmm… maybe like if I would change it like, there's different body type before some people might work out and some people like me just like exercise some times, and um there's like fat and not fat and or strong and an OK build, that's what I would say […] I mean, for me I feel like as long as I'm healthy and I, that's fine that's all I really want to ask for not like 6 pack [laughs] no no, I can't do that, I mean if I have that I don't look like a game developer [laughs]

(a) Example male avatar (b) Example female avatar

Figure 5.2: Example avatars from Kinect Sports Rivals

The word ‘sexualised’ was used by Miranda and alluded to by Nastasha who commented on breast size. This critique of being objectified was not discussed in these terms by men, who saw the body more as a cultural norm, rather than a sexual object. Again, KSR has restricted the appearance of the avatar to a set of rhetorical norms.

Alistair saw the avatar as an idealised (rather than ideal) version of himself, and this broke the illusion of realism that some of the other participants experienced:

> ALISTAIR: I’m not sure, it’s a little bit, with this avatar, like it’s not a very accurate, description is it? like in the sense that of course it wouldn’t know, I’m fully clothed, in black, so it’s it’s, now the idealised of.. what a male body should look like in a way.
>
> OC: do you think it’s what it should look like?

> ALISTAIR: Look at the legs [laughs] and arms, it's it's what it should, I don't think there's such a thing as it should be like this or that, it's a, fit version of a male body so it's quite athletic, quite all these sorts of things.

To some extent, the responses to the avatar are consistent with Virtual Self Discrepancy (Jin, 2012) who found that a greater discrepancy between the self and the avatar was related to a reduction in self-presence. Indeed, the work by Mancini & Sibilla (2017) notes that player-avatar identification[5] is reduced with wider self-discrepancies. Although the avatars were equally idealised across participants, body image and subjective norms may have influenced the perceived discrepancy between avatar and self, an avenue that could be fruitful for future research.

A further point to note is the critique that some participants voiced over the avatars. They actively engaged with the content they were presented and questioned the veracity of it in a number of cases, particularly when a greater discrepancy was perceived. This active engagement has been reported in previous studies on adolescents and adults discussing their relationship with the mass media (Cole & Grogan, 2019; Gill, 2012). This further supports criticisms of the 'Media Equation', or audience-as-passive vessel approaches to media influence studies, in which observers are merely static elements of a calculation, rather than constructing and perceiving beings. Whether this is representative of peoples normal media consumption is of course unknown, since participants knew they would be interviewed, and may have been more critical of the content than they would otherwise have been.

Theme 3: A foot in the door

The extent to which participants believed the game could be a substitute for exercise was discussed. An overwhelming majority of participants dismissed

[5]The extent to which a player feels they share goals and experiences with an avatar.

the usefulness of the exercise game in their own routine, claiming that it was not a viable alternative to exercise for themselves. Some participants had used exergames in the past [Keira, Selphie, Nastasha], although these were structured exercise programmes with gamified elements, rather than sports games with an exercise component. Cortez noted that the game was not as tiring as 'real tennis'. In fact, only two of the 20 participants claimed that they would rather play an exercise game (Alistair for privacy, Cait for fun).

> CAIT: I'd rather play a game like this than go to the gym, 100%, errrm because it makes it fun doesn't it, because you're doing this and maybe also it's not quite as far as going to the gym, but at the same time it's more fun isn't it, that's the main reason that I joined this society, because there's a reason, a fun reason to go other than just exercise.

> ALISTAIR it seems this is something I could do, like one of the things I don't like about exercise is. . . being seen by other people just because I'm not, I'm not very fit I'm not very accomplished, I go to the gym and just don't know what things do or don't, so it's easier if you was someone like that where the only things that can see you is just a bunch of pixels and polygons, it's fine, who cares?

Although just a small subset of participants disclosed this, most believed that people who would not feel comfortable at the gym or were maybe unable access traditional exercise methods may benefit from exercise games. Indeed, one participant [Aaron] said that that it may have potential for the future if his mobility deteriorated. The reasons for preference of exergaming over 'traditional exercise' voiced by Alistair and Cait are supported by health psychology literature. Fear of judgement and physique-related anxiety are frequently reported as barriers to exercise (Crawford & Eklund, 1994; Frederick & Morrison, 1996; Ross & Melzer, 2016); as is boredom (Miller & Miller, 2010; Pridgeon & Grogan, 2012).

Interestingly, the people who felt exergames were a good pathway to exercise were amongst the least active of the participants, suggesting that targeting inactive individuals with exergames could have potential. However, it seemed imperative that people get to choose their own ideals, and were given enough freedom, through affordances, to represent themselves. Meryl likened the experience to the body positivity movement surrounding a larger bodied man in Los Angeles who was filmed dancing. She stated that being able to see a larger bodied person exercising might have a positive effect; this may be the case, but what Meryl did not know was that this affordance is not available in KSR. In contrast, Anwen thought that a larger person seeing themselves as a larger avatar may be demotivational! Making avatars over-muscular or sexualised may break the feeling of self-presence.

Indeed, as was mentioned earlier, this follows from the work of Jin and Mancini who suggested that virtual self-discrepancy may reduce the feeling of self-presence or identification with the avatar (Jin, 2012; Mancini & Sibilla, 2017). Work on these two constructs has been specifically conducted with reference to exergaming. Li & Lwin (2016) ran a month-long intervention on school children in which they played an exergame. They found that the children who identified more with their avatar reported more self-presence and identification, which mediated enjoyment, exergame intentions and exercise intentions. By these models combined (*eae sunt* VSD, exergame motivation model), over-idealisation of avatars, or restricting affordances to a limited set may have the opposite effect to the one suggested by Li et al. (2014) and players may be less likely to exercise using a videogame.

Theme 4: Societal Norms and Subjective Attitudes: The Manifestation of Weight Stigma

The *fat is bad* rhetoric that was prevalent in the KSR game, and previous literature on avatar body types, was also manifest in the discussion of bodies. Primarily this was in the form of social norms, in participant's accounts surrounding their own bodies, and in their approaches to creating virtual bodies. Conformity to, or at least awareness of societal norms and the social undesirability of larger bodies was reflected in many of the accounts, directly and indirectly; as was the conflation of weight and health. However, some participants were also aware of these societal pressures and discussed them critically. Ward described the internal discussion between the norms and his own values:

> WARD I sometimes think oh maybe you could look a bit more toned, if you know what I mean, but then at the same time in a really cheesy way I almost see that as selling out like wanting to be toned, I think as long as you are healthy that's the main thing.

Some people specifically did not want to be fat in the game, and also thought that larger bodied people would be put off by seeing themselves with larger bodies. The expectation of shame is on brand for (at least) western society. They did not want to be fat and assumed that no one else wanted to be either. However, some thought that representations of larger bodied people in the media could be inspiring in spite of the size and could increase self-efficacy. Meryl discussed how an internet meme led to positive reception on Twitter, and that the presentation of larger bodied people in videogames may be inspiring:

> MERYL: Basically it's just this bigger guy having the time of his life dancing, and it he got a lot of hatred, and he was so upset but then loads of people reached out to him and they bought him to LA and they had fat man party but it's to show the fact that if you are bigger it doesn't necessarily know you are unhealthy and you can still do as much dancing as you like or whatever, it really made me happy but I suppose the issue is I think that skinny doesn't necessarily mean healthy but I suppose quite a lot of people would assume so maybe if for example it would help if for example if someone was bigger and

they see someone who's big on that having a great time it would be like well 'you can do it I can do it'.

The work of Pearl and Puhl has suggested that reducing the negative portrayals of larger bodied individuals may reduce weight stigma (Pearl et al., 2015b). However, Pearl et al. (2015b) only demonstrated this in *neutral* settings such as in office environments, rather than exercise settings such as the gym. That is, portraying larger people exercising may have negative connotations, although much further work is required to confirm this.

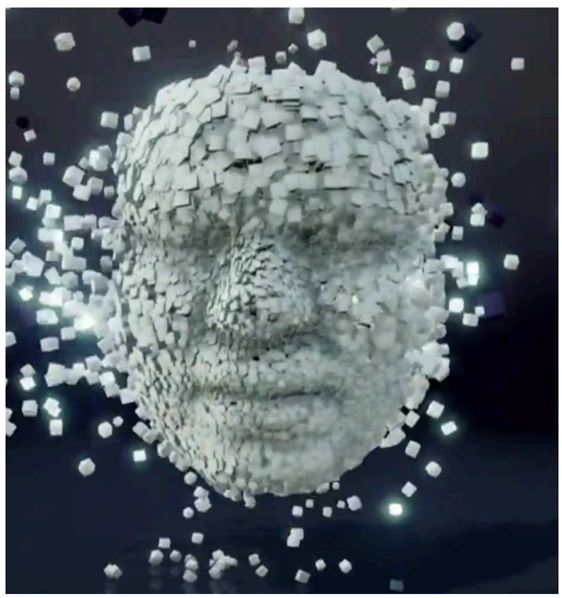


Figure 5.3: Cube-formed baby face described by Ward.

Not being portrayed as fat in game was important to participants. They voiced relief at the avatar not being fat[6], despite other inaccuracies such as 'skinny legs' or 'large hands', as one participant said "I'd say it's it's pretty good though, well it's not fat I suppose". Another participant claimed, "Well I think if she came up like really fat and looked unhealthy then I wouldn't like it, but she looks like quite healthy so, lovely avatar". This motivation to not be fat in game was reflected in participant's own avatar creation, their reactions to the avatar at various points (in response to the baby faced avatar, "Do I have cheeks that

[6]Pseudonyms are not provided here in case the interpretation is deemed accusatory.

fat?" Figure 5.3), and can also be seen in the wider gaming community. Indeed, one of the least used characters in the fighting game Tekken 6 is called Bob (Figure 5.4), who is a 'morbidly obese' fighter with the same speed and range of techniques of the thinner characters. The character is relatively easy to play and is discussed on forums as a 'good starter player'. However, he ranked with other marginalised groups such as a woman of colour (M.Raven), a homosexual male (Lee), and novelty characters such and bears (Panda and Kuma) as the least used characters in online tournaments.[7] To some extent, Bob is an example of how larger bodied individuals can be portrayed positively in the media, but there are also various stigmatising elements about him. For instance, all of his techniques are named after chocolates, or other allusions to digestion and eating ('egg bound'). According to the Tekken Wikipedia page, Bob was ranked the tenth "ugliest game character" in 2008 with the review "If the bright red shirt covering the gigantic frame isn't insulting enough, he also has a bleach blonde mop on his head"[8]. In order for people to be happier with larger bodied representations, these need to be more diffuse, and more widely available. However, for these representations to be more diffuse, people need to be more accepting of larger bodies: a veritable Catch 22.

Results Summary

In sum, many of the people interviewed in this study created avatars that portrayed their subjective ideals. These avatars either directly reflected the self, or represented an entirely different form that they would ideally like to be. Although these subjective ideals often involved being thinner or more muscular than the player, the *prescribed ideal* avatars generated by KSR were seen as a step too far by some participants, and regarded as idealised, hyper-sexual, or homogeneous. There were underlying anti-fat, or fat-phobic attitudes evident in

[7] https://www.neogaf.com/threads/tekken-ps4-pc-character-usage.1462134/

[8] https://en.wikipedia.org/wiki/List_of_Tekken_characters#Bob_Richards/Slim_Bob/Summer_Bob

Figure 5.4: Bob Richards from Tekken 6 and 7

the accounts provided by participants. Exergaming was widely believed to be a viable alternative to traditional exercise, only two participants claimed that they would be interested in adopting it themselves.

5.5 General Discussion

People varied considerably in their approaches to avatar design; this reflected, in part, the Relational and socio-constructivist approaches summarised by Mancini in that some participants retained their actual self in game, and others explored different forms; but there was also evidence that avatar creation could vary within participants depending on the purpose of the game, and the process could be seen as a game-space in and of itself. The goals were often to represent a subjective ideal, but the avatar was sometimes seen as a means to an end (avatar as object or game-piece) (J. Banks & Bowman, 2016b). Despite reportedly creating their ideal selves in game, the prescribed 'athletic' ideal that KSR generated based on the participants general feature was unwelcome in some cases; it was seen as either too sexualised, too muscular, or as not muscular enough. Participants discussed the abnormal body shapes and suggested that a greater range of body types would have been preferable. Although only two of the participants believed that the exergame might have a positive influence on their exercise behaviour, the majority felt that it could be a good foot in the door for people who were less active. The two people who though the game would be beneficial to themselves discussed barriers to exercise such as physique anxiety and boredom as reasons for them not being more active. The minor tweaks that participants made to their own personal avatars appear to reduce the discrepancy between their ideal and actual selves, however being portrayed by the computer as a 'hyper-ideal' may have the reverse effect and result in a negative experience. That participants expressed relief at not being portrayed as 'fat' may suggest a 'sweet spot' at which an avatar is at its most motivating, but it is perhaps unlikely that a computer algorithm could determine this and may be best left to the individual player to decide.

It is interesting to note that the avatars that participants discussed creating for themselves did not sound similar in terms of physique to the one generated by KSR. Although Cortez discussed a triangular shaped body as part of his ideal, his avatar design choices did not match this. The changes were sometimes more subtle, and sometimes far more extreme. Although players were relieved at not having a 'fat avatar', the dichotomous world described by Li et al (2014) where avatars could be either fat or thin would not benefit these gamers either.

Relation of Data to the Proteus Effect

The Proteus Effect predicts that people change their behaviours in line with how they think other people expect their avatar to behave. Kinect Sports Rivals appears to be attempting a different form of manipulation, one that promotes self-efficacy by showing a player their ideal future self. Such a mechanism is more in line with the work on *Virtual Self Modelling* by Jesse Fox and Jeremy Bailenson, who showed that the number of repetitions of conditioning exercises could be increased by showing an immediate change in body-size in a virtual-self model (Fox & Bailenson, 2009). It was evident from the accounts that weight stigma influenced avatar choices. It is not the purpose of this study to generalise, but it does suggest that the general players of games like Virtua Tennis or EA Sports Active (used by Peña et al. (2016) and Li et al. (2014) respectively) are unlikely to select an avatar similar to the ones in the 'overweight' conditions of the Li and Peña studies, and so players are unlikely to experience the Proteus Effect-related 'slow down' that has been associated with these avatars. However, it does not help that these authors attached a *pseudo-health warning* to larger bodied avatars on the grounds of three quite heterogeneous studies!

The argument against this *health warning* forms the grounds of the rationale behind the upcoming studies in this thesis,including an exploration of the stereotypes associated with larger bodies and athletic bodies. It also aims to determine whether the Proteus Effect, as applied to larger and athletic bodies, will replicate. It will also explore whether providing additional positive information about larger bodied avatars will moderate the effect, if it does replicate.

There were some allusions to a Proteus-type effect in this study's accounts - for instance, Nastasha claimed that she felt the power in her avatar's legs was the power in her legs; but there were several in which the participants referred to the avatar as an object or an other (cf. Bank's and Bowman's player-avatar relations, page 22). To The Authors knowledge, there has been no adequate phenomenological investigation of the Proteus Effect, which would be an interesting future direction for qualitatively-inclined cyberpsychologists.

Reflexive Analysis

Every attempt has been made to reflect the accounts of the participants accurately and to interpret their responses sensitively. The topics were relevant to The Author's academic interests of virtual reality and body image as well as personal interests of videogames and exercise.

It is possible, if not likely, that some participants will have reacted differently to the experience and the content of the divulged information, particularly on sensitive issues such as body image, and would be different if the researcher were younger or a different gender.

In playing a game with participants which had competitive elements to it, the PhD candidate influenced the dynamics of the situation, placing himself and the participants as rivals which may have affected the post-play interviews. Wherever possible he tried not excessively beat or embarrass the participants and did not play to win, although in some cases where the participant was struggling on the climbing it made more sense for him to just finish the level and coach them.

The interviewer also detected that some reports on the game may have been influenced by his presence and believed that many of the participants may have thought that he had made the game or designed it in part. This is because most of the compliments were followed by some form of criticism about the game, and many of the participants seemed quite ambivalent towards it despite being verbally positive about it.

Strengths and Limitations

The study benefited from having a relatively loose inclusion criteria for the definition of gamer. This meant that a broad range of knowledge and experience could be sampled. It was also a long and varied interview, sometimes exceeding an hour, meaning that participants were able to share their experience and respond to the stimulus at length. Limitations to the study were the fact that the sample was almost entirely (N = 19) limited to individual from western countries (Germany, UK, Greece, Poland and Sweden) and was predominantly (N = 17) white. This sample did not therefore fully represent the diverse range of ethnicities within the city of Manchester and the UK in general. The failure to adequately sample from different ethnic groups means that possible cultural differences in gaming, attitudes to exercise and body image could not be identified and included in future studies and possible interventions may be made available within the communities that these individuals live in. Some issues regarding the representation of race were raised, with one participant saying that the avatar ”does not look very Chinese”. Where other participants mentioned that the face was not representative of themselves, they did not discuss this in terms of race. The absence of sound was also a limitation. Some participants wanted to hear the audio feedback from the tennis ball hitting the virtual racquet, and others made comments about it being quite strange without sound. There were also technical problems with some of the interviews in that the Kinect did not respond well to black trousers and lighting conditions made the avatars hair appear darker. Where possible these were fixed, but it also meant that the quality of the experience of the participants varied slightly

Applied Context

The findings from this study may be used to guide avatar design towards providing a more inclusive array of affordances and eschewing prescribed ideals. The applied directions from previous, purely quantitative studies have included the restriction of avatar body types, where the purpose is promoting exercise, to ’normal’ to athletic physiques. This study has responded to these claims and provides updated guidance that restricted ranges of avatar body types might be off-putting to some people. The study also lends some qualitative support to the idea that exergames may offer an alternative form of exercise to people who have psychological barriers to exercise (e.g. Alistair’s physique and accomplishment anxiety), and that an increase in the range and accessibility of such games might be beneficial. The key finding in this study is that players need freedom to choose their avatar, and that sufficient options should be available to allow the either flexibility to be as they are, rather than having their choices restricted by the game. For those who do not care about the visual appearance of their avatar (those who have an *avatar as object* relation with the avatar, page 22), a range of defaults that do not conform to normative ideals might help normalise a wider range of body types. This contrasts with the suggestions made by the authors of some of the papers in Chapter 4 that when exercise is the goal, bodies should be restricted to athletic or 'normal' weight avatars.

Findings and Conclusions

The purpose of this chapter was to respond to the research discussed in Chapter 4 which provided suggestions for the range of affordances available to users during exergaming by exploring gamers' accounts of presenting themselves in digital environments, and their responses to being portrayed in an idealised way. This study also provides an insight into the aversion of larger bodied avatars in everyday gaming, which appears to be related to fat-phobia and weight stigma as discussed in Chapter 2 (page 44) It is interesting to note that larger bodies were conflated with poor health, and the gamers in this study did not discuss larger bodied people in terms of laziness - as if this conflation is an acceptable stigma. Evidence of negative attitudes towards athletic people (Chapter 2, page 47) was largely absent from the gamers accounts, although in some cases the bodies were not desirable to the participants, who preferred to determine their own ideals. The choice of an ideal represents an active engagement with visual media, such as avatars, which is distinct from simply being exposed to idealised media (see Chapter 2, page 44). As was discussed earlier, the aim of exergaming is to make a sedentary pastime active (page 17), and that the gaming population may benefit from this. However, in the accounts from this study, many of the gamers reported being quite active. In only two cases would people rather play an exergame than engage in traditional exercise. This would suggest that the target population for exergames is those for whom exergames are a necessary component of exercise, which is a much narrower population.

Conclusion

I have attempted to show that some participants aligned themselves with relational, and some a socio-constructivist approach to avatar creation, and that this may in part be informed by subjective ideals. I argued that, although being presented as an ideal is sometimes welcome, nuance is needed if and when digital representations of players are to be restricted to western socio-cultural body norms such as muscularity and slenderness. People discussed making minor changes to their bodies or creating entirely new characters based on their ideals. Having control over these parameters seemed to be key to the effectiveness. Indeed, Ness suggested that were he unable to reflect himself in-game as either realistic or a cartoony, he would opt for an identity tourism approach, roleplaying a different character altogether. Restricting someone with these preferences to an athletic version of themselves may be counterproductive when trying to inspire people to play exercise games for health. Negative attitudes towards larger bodies were prevalent in the accounts and may suggest that larger bodied characters would not be used in videogame, even if available. One of the aims of this thesis is to determine whether representations of larger-bodied characters in games may be improved by providing additional information and challenging the *self-evident truth* that larger bodies are unfit. Much of the former research on the Proteus Effect has relied on these *visual aphorisms*, and one of the goals of the thesis is to investigate whether these aphorisms may be interrupted with salient behavioural information.

Chapter 6

Lean and Mean: Exploring stereotypes associated with athletic bodies

6.1 Rationale

Chapter 4 and Chapter 5 covered many complex topics and raised several questions. Following on from the conflicting findings between the assertions by Li et al (2014) discussed in Chapter 4, and the opinions of participants concerning muscular and idealised avatars in Chapter 5, the remaining chapters will thoroughly investigate the implied but unconfirmed stereotypes relating to athletic bodies, and a deeper investigation into th negative steroetypes associated with larger bodied people. These stereotypes are necessary for a Proteus-related behavioural response to be evoked (see page 52) and addressing the implications of Based on the prior literature (Chapter 4), the outcome of interest for this thesis was determined to be *exercise intensity*; the intervention was the provision of *athletic vs plus-sized* avatars. Information about the stereotype space that these representations occupy is incomplete. It is possible that the previous examples of the Proteus Effect in physical activity research were based upon the negative stereotypes associated with larger bodies, or positive stereotypes related to athletic bodies (or, indeed vice versa). It is also unknown whether athletic stereotypes are exclusively positive, and plus-sized stereotypes exclusively negative. If the proposed *dumb jock* stigma for college athletes also applies more generally to people with athletic body types, or the *smug* jock discussed by one of the participants in Chapter 5 is more general, this may have consequences for future interventions (page 47). In this chapter, athletic and plus-sized *prototypes* will be dissected, and a thorough investigation of the stereotype content space that they evoke will be established. This data will inform the later studies in this thesis.

6.2 Introduction

There is evidence that stigmatising attitudes towards plus-sized individuals are held by members of general population (Puhl & Brownell, 2006; R. Puhl & Suh, 2015), people who exercise regularly (Flint & Reale, 2018), and fitness professionals (Ntoumanis et al., 2018). Perhaps as concerning (if not more so) is the observation that this stigma may be internalised by people with larger bodies (Tiggemann & Rothblum, 1988). In contrast, there appears to be an assumption that people who exercise regularly are *motivational* or *inspirational* for people with body types ranging from normative-ideal to plus-sized body types. Despite assertions that exercisers are revered, such as those made by Ginis & Bassett (2012) in the *Encyclopaedia of Body Image and Human Appearance* (full quote on page 47 of this thesis), there is no supporting evidence to suggest such favourable stereotypes. Indeed, the majority of research on this topic suggests that athletic individuals are negatively evaluated (see section 2.5, from page 47).

For example, there has been some exploration of sports-related stereotypes, such as the 'dumb jock' stereotype (Sailes, 1993); gender 'appropriate' sports (Chalabaev, Sarrazin, Fontayne, Boiché, & Clément-Guillotin, 2013); and racial stereotypes, with white athletes being internally represented as 'hard working', and black athletes described as 'natural sportspeople' (Moskowitz & Carter, 2018). In only one Implicit Association Task study were athletic stereotypes conceptualised as the diametric opposite to the fat-related 'lazy' stereotype (Robertson & Vohora, 2008). This is an important point because it suggests that the assumption that the two physiques are diametric opposites is unsubstantiated.

However, there is a complex relationship between exposure to athletic bodies and outcomes such as exercise performance and appearance-related affect. In a study into body-image and Social Comparison Theory (Festinger, 1954), it was shown that upward comparisons with 'superior others' may result in lower body satisfaction (Robinson et al., 2017). Elsewhere in a naturalistic experiment, Wasilenko et al. (2007) demonstrated that participants who were unwittingly paired with a fit-looking peer during exercise, exercised for less time, and had lower body satisfaction compared with those who exercised with a larger-looking peer. In the qualitative literature, unfavourable comparison of the self to the bodies of others was reported as a reason for gym non-adherence by women in an interview study by Pridgeon & Grogan (2012). However, upward social comparison with 'a guy like that', (meaning the person who could lift the most weight), was also inspirational for some men in Pridgeon and Grogan's study. Halliwell et al. (2007) found that exposure to media images of muscularity increased negative body affect in men who did not exercise but had the reverse trend in men who did exercise. Conversely, exposure to 'average' sized models has been associated with an improvement in body-related affect (Diedrichs & Lee, 2010, 2011). Finally, Pila et al. (2016) demonstrated that men with negative body evaluations participated in more exercise when they made body related upward social comparisons, whereas the reverse was observed in women. In sum, there may be negative and positive aspects to the use of 'athletic' others as motivational stimuli.

The purpose of this exploratory study is to estimate the stereotype structure related to body types, specifically athletic, compared with plus-sized body types. To do this, participants will be provided with only text information about gender and body type, that is, there will be no indication of occupation, hobbies, or race[1]. The importance of this study stems from the pervasive assumption that the general public regard the bodies of athletic individuals as desirable, resulting in a positive 'upwards' social comparison (Collins, 1996). This assumption informs promotional materials, advertising, and body-related affordances of media representations. However, there may also be negative elements to this representation, for instance athletic body-types may evoke evaluations of arrogance or being judgemental.

Measuring Stereotypes

Early examples of stereotype research include the 'checklist' method by Katz & Braly (1933), who instructed a small group of undergraduate students to generate words associated with different racial and ethnic groups. A larger group of undergraduates then completed a checklist with these words, indicating whether or not they felt the word was applicable to the group. Another method of stereotype measurement is the percentage approach. This involves instructing respondents to estimate the percentage of people in a group who have a particular attribute. Larger percentages (perceived proportions of the group population) indicate stronger stereotypical traits (Brigham, 1971). In addition to the percentage of traits in a given group, further information may be acquired by asking participants to estimate the prevalence in the general population. By dividing the group by the population percentage, a ratio of diagnosticity may be calculated. A larger ratio indicates a stronger stereotypical trait (Martin, 1987).

Later examples include the prototype and pathfinder methods. Under the prototype technique, participants are instructed to list traits or characteristics that they believe are associated with a 'typical' member of the group (Cantor &

[1] In the word generation phase, participants were exposed to either text or visual representations of the groups.

Mischel, 1977; Rosch, 1999). The pathfinder method is a network model whereby a diagram is created by plotting characteristics based on how closely related or similar respondents rate items as relevant to a label. In a measurement review using stereotypes of Russians and Americans, Stephan et al. (1993) report that each method accesses similar cognitive processes. The adjective checklist approach has been used recently to identify stereotypes associated with women in politics (Schneider & Bos, 2014), and modern groups such as online gamers (Kowert et al., 2012). These methods have been adapted to include more scale points, allowing a more granular representation of the latent stereotypical construct (Cuddy et al., 2009; Kowert et al., 2012)

In the current study, the structure of stereotypes related to body-type was explored using the adaptation of the checklist method used by Kowert et al. (2012). This was because this method allowed for both the latent structure of the stereotypes to be explored, as well as the strength of the perceived applicability of the stereotypes to different groups. Participants were provided with the description of a group, rather than a particular individual. The study has three methodological phases:

1. Stimulus validation: A set of avatars were created and evaluated by body image experts.
2. Word generation: Participants generate words that they feel are stereotypically associated with descriptions or images of larger bodied and athletic individuals.
3. Word rating: A different group of participants rate the most frequent words in terms of applicability to text descriptions of larger bodied or athletic individuals.

.

6.3 Methods

Phase 1: Avatar Development

Avatars were developed during a structured consultation with an expert in body image (second supervisor) who has published extensively in the body image field. The routine for this consultation can be found in the digital appendices (https://osf.io/a37pc/).

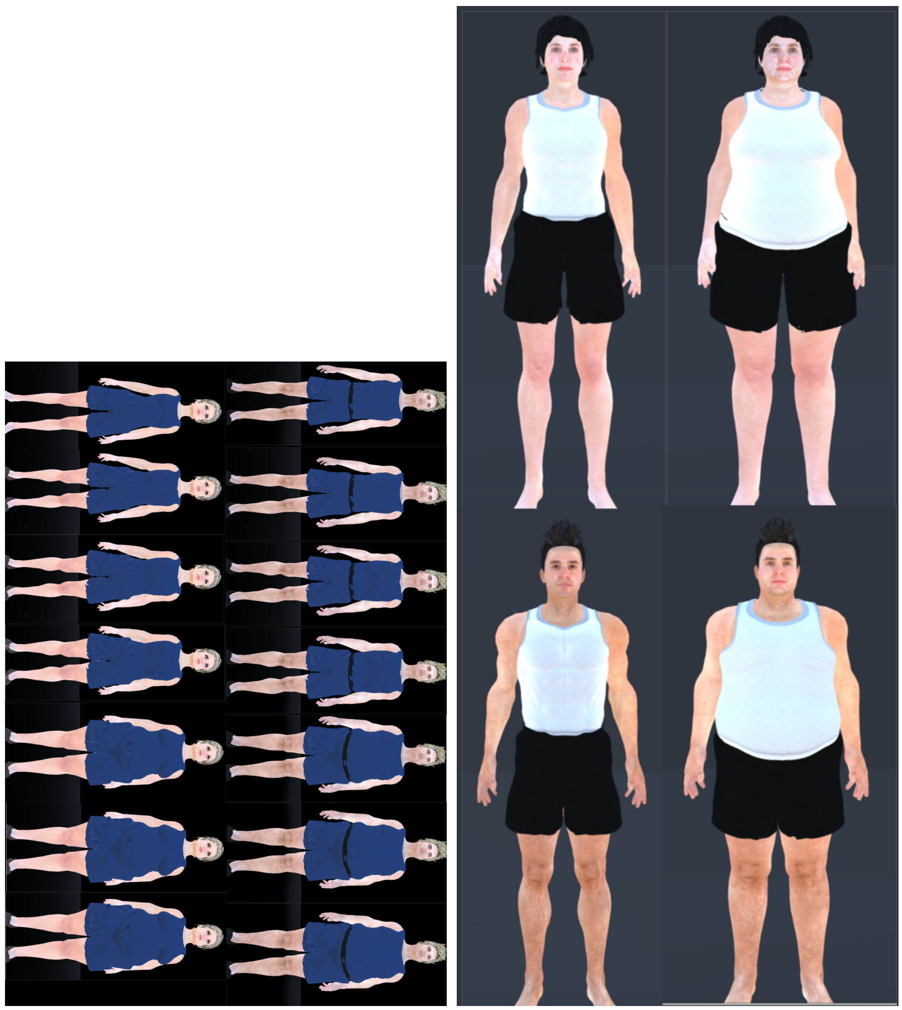
During this consultation, the criteria for selecting body image experts as avatar raters were discussed. It was decided that a benchmark of 10 body-image related scholarly papers, regardless of metrics such as impact factor would be the inclusion criteria for expertise. It was also decided that the raters should not be related to the programme, and so were excluded based on involvement in the project either at a supervisory or examiner level. The structure of the avatar creation process was separated into 5 segments, detailed below:

Apparatus

To design the avatars, Daz Studio 3D suite was used. A range of customisable assets are available for this software. Many of these are available free of charge with limited options.

Daz Studio 3D includes a selection of models for which a number of add-ons are available, including the Genesis range of customisable avatars. There are 8 versions of the Genesis avatar, and they are available in male and female sexes. During the consultation, the options available for the most accessible models (Genesis 2 and Genesis 8) were assessed, and the two versions were compared. Genesis 2 is an older version of the avatar, however there were more clothing assets available for it. Further, many of the advanced additional features were discounted from the online store, so more customisation options were available, given the budget. As such, it was decided that the Genesis 2 model would be used as a base model.

The base model was fitted with the free assets ‘Muscle Top’ and ‘Boxer shorts’. The default colour for these assets is black and so they were changed to Navy blue so that the avatar’s muscle textures were more visible through the clothes. They were also fitted with ‘WildeMane Hair’ and ‘Toulouse Hair’ texture for male and females respectively.



(a) Landscape from left to right: Base avatar shape; three athletic shapes; three plus-sized shapes

(b) Portrait top: Female athletic and plus-sized avatar; bottom: Male athletic and plus-sized avatars

Figure 6.1: Evolution of the avatar design process

Preliminary decisions

There was also a brief discussion about what features an athletic individual would have regarding muscle and fat ratios. Athleticism was defined in this case as an all-round fitness, rather than dedication to a particular sport. As such, a prototypical athletic individual was defined as someone who goes to the gym very regularly, and/or instructs people in exercise. The V shaped physique was discussed in terms of its diagnosticity for athleticism in both men and women. This includes broad shoulders and a thin waistline. Evidence of low levels of body fat and the presence of lean muscle were also deemed to be important in the depiction of athletic individuals. It was also decided a priori that body building was an extreme example of muscularity and represented a different population of fitness enthusiasts and so hypertrophic muscles were decided against.

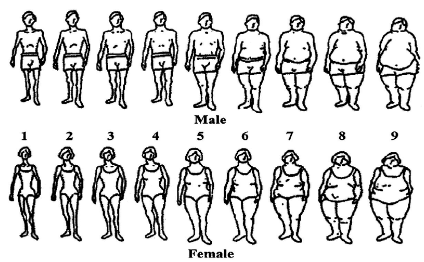


Figure 6.2: Stunkard Figure rating scale

Method

The range of each relevant parameter was explored for each avatar by adjusting the value from the minimum to the maximum, so that the boundaries were clear to the avatar designers. 'Micro' (individual body parts) and 'macro' (diffuse changes) level morphs were available in the Genesis 2 package. This means that avatars could be adjusted on a general (macro) attribute such as 'Fitness' or 'Portliness'. Making changes to these parameters changes the micro level morph points together, so that changes are distributed across the whole figure. For example, the 'weight' parameter changes all parameters (e.g., waist, arm, and leg circumference) simultaneously, such that the weight is evenly distributed across the figure. During the validation phase only macro-level changes were made to the avatars.

Once the parameters had been explored sufficiently, the most effective for portraying the different body types were selected. The specific parameters will be discussed below. In all cases the values were set, and the figure was saved. Three versions of each avatar were created, the one designed by the research team, and a larger and smaller version.

Avatar Development

All parameters discussed below scale from -1 to +1. Examples of the avatars can be seen in Figure 6.1a.

1. Female Athletic Avatars

To create the female athletic avatar the following parameters were set to a maximum of 1: 'body tone'; 'fitness details'; 'fitness size'; 'thin'. 'Weight' was set to the minimum of -1. To reduce the size of the hips, the 'androgyny' parameter was set to 1. 'Body size' was set to 0.5. This created a lean, muscular avatar. The larger and smaller avatars were created by changing the body size to 1 and 0.0 (mid-point) respectively.

2. Female Heavier Weight Avatars

To create a heavier weight female avatar, all parameters were reset to zero and the 'weight' parameter was set to 0.9 for the first heavier weight avatar. This corresponded to a Strunkard Figure Rating physique of 8. By changing the 'weight' parameter to 1.0 and 0.8, avatars approximating Strunkard figures 9 and 7 were developed.

3. Male Athletic Avatars

To create the male athletic avatar the following parameters were set to their maximum of 1: 'body size'; 'body tone'; fitness details'; 'fitness size'; and 'thin'. 'Weight' was set to the minimum of -1. Again, the 'androgyny' parameter was set to 1. This created a lean, muscular male avatar. Unlike the female athletic avatar, setting the 'body size' to 0.5 in the male avatar made it look too thin, so it was kept at 1. This meant that an extra parameter was needed to make the larger athletic avatar. The 'body builder size' parameter was increased to 0.3, which was the smallest change that the team could detect. To make the smaller athletic avatar, 'body Size' was reduced to 0.5.

4. Male Heavier Weight Avatars

To create a heavier weight male avatar, all parameters were reset to zero and the 'weight' parameter was set to 0.9 for the default heavier weight avatar. This corresponded to a Strunkard Figure Rating physique of 8. By changing the weight parameter to 1.0 and 0.8, avatars approximating Strunkard figures 9 and 7 were developed.

Round 1 Validation

A group of experts in body image (N = 10) were approached by email and asked to fill out a short questionnaire about the original 16 avatars. Expertise was defined as a researcher who has published at least 10 body-image related scholarly papers, regardless of metrics such as impact factor. Of the 10 researchers approached, 7 responded, and two were unable to access the images so they were asked to rate the second version of the exemplars. The questionnaire involved using a slide bar to indicate the amount of muscle and weight that would make the avatar look maximally athletic or sedentary. Two open ended questions were also included in the survey for people to include a qualitative critique of the figures. The survey was presented using Qualtrics (Snow & Mann, 2013) and results were exported to CSV file to be interpreted by the researcher.

Three main points were made by the respondents. The first was that weight influenced their decisions in the athletic avatars, rather than muscle which was hard to detect in the images. This was the same for the larger bodied avatars. That is, the presence of adipose tissue was a stronger cue than the lack of muscle. A second point was that although the athletic set was smaller, the lack of muscle definition made them look less fit. Finally, some of the male avatars had a 'feminine' appearance, which was mentioned by the respondents and the supervisory team. Some mentioned that the outfit that the avatar was wearing may have had a feminine look to it.

Second Attempt

The primary pattern in the numerical responses to the first set of avatars was that more muscle, and less weight was needed in the athletic individuals. Only more weight was needed in the sedentary individuals. In the second attempt this feedback was incorporated, as well as the comments that the male avatars looked feminine. To address the latter issue, the FaceGen Artist software was used to create a generic white male face model, with averaged features. Short facial hair was added to the face. This was exported to as a texture for the Genesis2 male avatar, and a new base male model was created in Daz Studio.

Once the texture had been added, the body morphs fitness and body builder details were increased to the maximum and the body builder size was adjusted to 100. For the sedentary avatar, a similar process was involved but with the body morphs 'heavy' increased to the maximum. The female avatars also had a FaceGen artist model created which was a white female with averaged features. Textures were added to the face to make the avatar look slightly more realistic. The faces and skin textures were added to a Genesis2 female avatar and morph shapes were adjusted.

This time the avatars given black hair and exported to the Unity software where the materials and textures were altered by including different shaders and adding normal maps. For all avatars, a specialist skin shader was used on the avatars. The shader has been designed so that developers can adjust how the texture interacts with light to better resemble skin (Cicada Studios). The overall effect was that the avatars appeared brighter and the surface details such as muscle tone and skin texture were more visible. This shader was also added to the MuscleTop clothing so that the abdominal muscles were more visible through the top. The avatars were sent to the supervisory team, and to the two respondents who could not see the images displayed in phase one for feedback. Only one respondent replied, but feedback was also received by the supervisory team. The avatars were deemed to be better but still feminine. It was pointed out that the pectoral muscles in the athletic male looked like breasts, the top was feminine, and the male sedentary avatar seemed to have breast tissue.

Third Attempt

In Daz Studio, the androgyny full body morph was adjusted to -0.5 (Maximum Male), which increased the size of the shoulders and reduced the chest. Further, chest scale and chest hang partial body morphs were adjusted so that the pectoral muscles seemed less obvious. Finally, a new top with a higher neckline was used (Tankini) and was made white rather than blue so that the contours of the abdominal muscles were more obvious in the athletic avatar. The avatars were edited in Unity with the same skin shaders as in the previous attempt. The final set of avatars can be seen in Figure 6.1b.

Phase 2: Word Generation Pre-study

Apparatus

After the final set of avatars had been agreed upon by the supervisory team, they were added to scenes in the Unity game development platform. The virtual room they were placed in had a charcoal grey background. The project was built and run on a Mid 2015 MacBook Pro with a retina screen using the highest graphics setting (Ultra) with the highest resolution (2880 X 1800). A screenshot was taken of each of the four scenes and the images were cropped to 912X1800 pixels. The final set of four models from the development phase of the study consisted a heavier weight and an athletic-muscular white male and female.

A short survey was created using Qualtrics and advertised on the social media platform Twitter, and by word of mouth.

Participants

The target sample size was 50, but 56 responses were eventually collected once incomplete responses were closed. No demographic information was taken. This number was chosen to be in line with previous research. Katz & Braly (1933) asked 25 students to generate adjectives words, whereas Schneider & Bos (2014) asked 42.

Method

In the pre-study, 56 participants provided free-text responses for either visual or word-based primes. In the verbal condition, participants provided traits and behaviours in responses to the following four descriptors: male athletic, male plus sized; female athletic, female plus sized. In the visual condition, responses were made to avatar exemplars of these descriptors. In addition, participants in both conditions were asked to generate words that were related to the people who are 'fit'. An 'unfit' condition was not included because fitness was of primary interest[2]. Participants always responded to the fit condition last.

[2] This would be an interesting inclusion in future replications.

Procedure

Participants were randomly allocated to one of two conditions using the Qualtrics block randomiser function (refer to page 86 for more details). In the discursive condition, participants were asked to:

Please list as many specific stereotypical traits and behaviours that you think are associated with [women $\vee$ men] who have [overweight or plus-sized] $\vee$ [athletic or fit] body types.

Participants in the visual condition were presented with the four avatar images and asked to list as many stereotypical traits and behaviours they felt were associated with people with body types like those depicted in the images. The order of the questions was random.

Participants in both conditions were then asked to respond to a final discursive stimulus, listing as many stereotypical traits and behaviours they thought were associated with people who exercise frequently. This was added to the end of the survey because exercise is a behaviour, and it may have influenced participants responses.

Participants were then given the opportunity to sign up for a prize draw from which they would win one of three £10 book vouchers.

Text Mining

A text mining approach was deemed more appropriate for the construction of the word list since this minimised the researchers interpretation at the point of filtering and provided a reproducible list, based only on word frequencies. Alternative approaches include manually counting the words, or Thematic Analysis in which latent structures of the texts were explored (Braun & Clarke, 2006), but priority was given to reproducibility. Data and scripts are available from the OSF page (https://osf.io/ngkd5/).

Preparation Responses from the Qualtrics word generation survey were downloaded as a Tab Separated Value (TSV) file. Visual checks were made for rows that had clearly not been completed correctly. One row was removed for only containing single letters. The full set was then separated into 9 separate files: one for each modality and (sex:body type: modality, fitness words) and the header rows were replaced with short titles. The text mining method was implemented using the tm package from R (Feinerer & Hornik, 2018; Feinerer, Hornik, & Meyer, 2008). Each file was sequentially imported into R and converted into UTF8 format. Newline characters were replaced with spaces, trait and behaviour columns were combined, and duplicated words from each row (participant) were removed so that certain words were not over represented within. The individual files were then converted to dataframes in preparation for pre-processing.

Pre-processing Each dataframe was converted to a `Corpus`, which is a data object that is readable by the tm package. All words were converted to lowercase; punctuation, numbers and stop-words (e.g., conjunctions, pronouns, adjectives) were removed; datatype formatting was removed; and words were stemmed by reducing them to their root form (e.g., lazy and laziness both become lazi).

Mining Each corpus was then converted into a text document matrix which contained each word and its frequency. Word clouds were then created for each of the 9 groups. Following previous research (Schneider & Bos, 2014, footnote, p251), any word that appeared within a dataset 3 or more times was included. Similar words were combined into higher order categories. Within these cases, only one word from each higher category was included. The final list from the text mining procedure was predominantly word stems. After duplicates from the entire set were removed, the stems were manually changed into lemmatised adjectives. Verbs, and nouns that suggest an activity were included (e.g., gym), but were presented as predicates in the habitual present tense (e.g., “goes to the gym"). After synonyms were removed and the words had been reconstructed, there were 53 in the main list.

Phase 3: Word Evaluation study

Design

The study had a 2X2 between-subjects design, with participants randomly assigned to one of four conditions (Man/Woman:Plus-sized/athletic). Randomisation was achieved using the Qualtrics survey platform (Snow & Mann, 2013) and the research team were blind to the order.

Procedure

Participants read the participant information sheet and gave their informed consent. They then answered some questions about their gender and age group. Next they were presented with a sentence describing a group of people. The descriptors were:

[Men $\vee$ women] who have ["overweight or plus sized" $\vee$ "athletic or fit"] body types.

They were presented with 53 words in a random order and asked to rate each one on its applicability to the target descriptor. Responses were on a 7-point Likert scale with two anchors ('Not applicable', 'Very applicable') at the two extremities. After participants had responded to all 53 words, they were thanked for their time and provided with a link to a survey that would allow them to have their data removed from the study if they provided an anonymised identifier. The Qualtrics survey file can be found on the project OSF page (https://osf.io/ngkd5). Participants received no financial reimbursement for participating, but Psychology students at Manchester Metropolitan University received 10 participation pool credits for finishing the study.

Participants

Inclusion Criteria

Participants had to be over the age of 18, have normal to corrected-normal eyesight, and be able to understand written English to participate in this study. The reason for the broad inclusion criteria, and not including restrictions for non-gamers, was that this study was concerned with participants knowledge of stereotypes associated with relatively common in-groups (larger bodied and athletic. It was assumed that no specific knowledge or experience was required to have an understanding of how these groups are perceived.

Sample Size

Two-hundred people participated in the study. These people were approached primarily through the Manchester Metropolitan University Research Participation Pool, however other methods included posts on social media (e.g., Twitter, Reddit), and emails sent out by the research team to students. Ideally, a minimum of 400 participants would have been collected (Goretzko, Pham, & Bühner, 2019), but due to time constraints, sampling was ceased at 200. Results are interpreted with this potential weakness in mind.

Exclusions

Data from 6 participants were excluded because their responses were incomplete, leaving 194 to be included in the analysis.

Demographics

Frequency data for age and gender are provided in table 6.1. It is important to note that the vast majority of participants were young and identified as female. This is likely due to the biased pool from which participants were approached,

Table 6.1: Age and gender of the participants in this study.

	Female	Male	Non-binary	Prefer not to say
Prefer not to say	1	3	0	1
18 - 24	97	23	1	1
25 - 34	15	15	0	2
35 - 44	13	4	1	0
45 - 54	11	1	0	0
55 - 64	1	0	0	0
65 - 74	4	0	0	0

rather than the appeal of the study itself. There are large gender gaps in students who study Psychology, with women far outnumbering men (see Figure 6.3), and so where this sample might be representative of psychology students, it is certainly not representative of the population of the UK. This may have biased the findings, and more balanced samples were sought in subsequent studies.

6.4 Results

Exploratory Factor Analysis

Responses to the verbal descriptors were explored using the R package `psych` (Revelle, 2018). The KMO() function was used to check the sampling adequacy of the dataset using the Little Jiffy index (Kaiser & Rice, 1974). This function calculates the sum of squares for the columns of the correlation matrix, divided by the sum of squares of the original correlations which provides a ratio of the amount of variance explained by the columns with the overall variance. This measure was run repeatedly after N = 100 until sampling adequacy was above 0.80. Next, the `fa.parallel()` function was used to explore the number of factors and components within the data. Parallel analysis compares the scree plots for the data with simulated data from an identity matrix of the same size (only zeroes, with ones on the diagonal). Data were ordinal, and so polychoric correlations

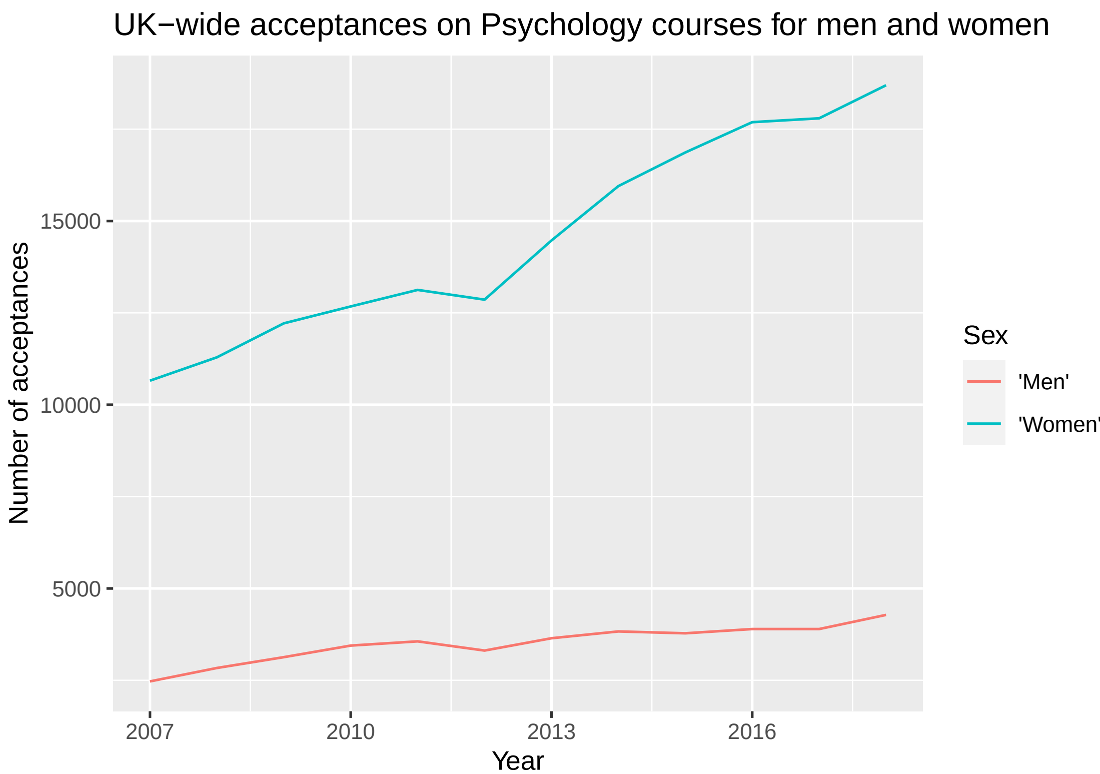
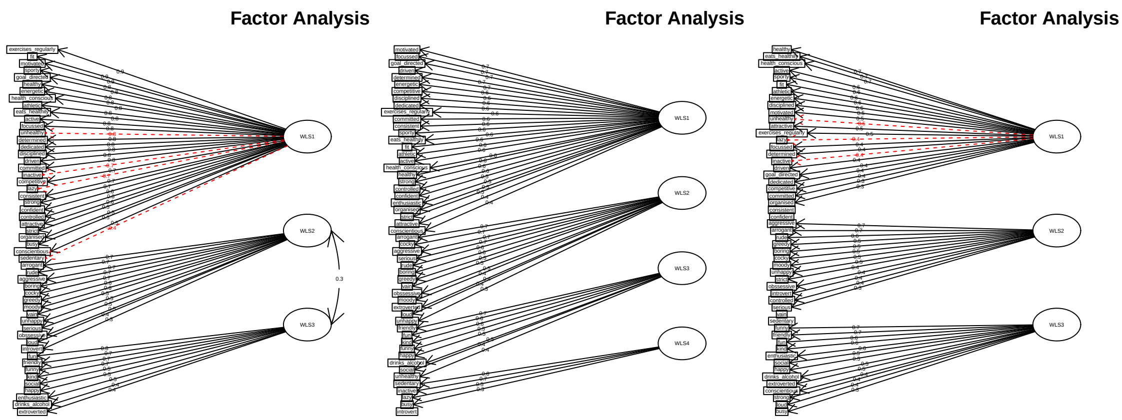


Figure 6.3: Plot of the number of UK Psychology degree acceptances between 2007-2018, ploted by the sex of the applicant. Source: UCAS 2018.

(rather than Pearson's r), and the Weighted Least Squares method of parameter estimation were used (Holgado–Tello et al., 2010). Because it was assumed that factors would be correlated since they were drawn from similar stimuli, the oblique rotation method `oblimin` was used, (Tabachnick & Fidell (2007); pp639-640). Three factors were reliably found both in the full dataset, and in the sub-sets of the data (i.e. plus-only, and athletic-only).
Once the number of factors had been identified, factor analysis was run using the `fa()` function in the psych package, again using polychoric correlations and weighted least squares. Factor diagrams for the full datasets and data subsets can be found in Figures **??**,**??**, and **??**.

Fit of the models were determined by the Root Mean Square Error of Approximation (RMSEA) index, which is an estimate of *close fit* of the factor (Flora & Flake, 2017) and Tucker-Lewis Index of factorability. The fit indices for each model are presented in Table 6.2



(a) Full dataset (b) Athletic only dataset (c) Plus sized only dataset

Figure 6.4: Factor diagrams of the three factor solution.

Table 6.2: RMSEA, SRMS, and TLI for each model

	RMSEA	lower	upper	confidence	RMSR	TLI
Full data set	0.04	0.02	NA	0.90	0.04	0.97
Fit filter	0.30	0.25	NA	0.90	0.06	-0.03
Plus-sized filter	0.29	0.25	NA	0.90	0.07	-0.03

Adequacy of model fit was judged using recommendations from Hu & Bentler (1999), who suggest that a combination of measures be used to manage type 1 and type 2 error rates. Since the sample size for this study was relatively small, and well below the suggested N = 400 (Goretzko et al., 2019), results should be interpreted with caution. Only the full dataset showed adequate model fit, with too few observations in the data subset (e.g. athletic or plus-sized only) models to be confident in the estimates.

Factor Descriptions

The Eigen values and factor loadings for the three factors are presented in Table 6.3.

The factors broadly represent health and competence, arrogance and vanity, and friendliness and sociability.

Health and Competence The first factor was loaded with items including *Dedication, Confidence, Eating Healthily, Exercising Regularly.* These items suggest that the latent structure that generated the words is one of self-control and efficacy. Weaker loadings included physical descriptors such as 'attractive', but these were not included in the final set. It was surprising that the health and competence items did not load onto separate factors, since it is possible to be dedicated to activities other than those related to health[3].

Arrogance and Vanity The second factor was loaded with predominantly negative traits, including *vain, cocky, aggressive*, suggesting that the latent construct that generated the representations was one of self-interest. This may

[3]This was the case even when a four-factor solution was forced. When this happened, the negative loaded items from factor one broke off into the fourth factor. When the data were accidentally treated as metric in an earlier analysis attempt, the factor structure was more complicated and included separate factors for dedication and health - however, as Van der Eijk & Rose (2015) note, there is a risk of over-factoring when Likert items are treated as metric, there are large numbers of items, and small numbers of participants, on which charges this original analysis was guilty of all three.

Table 6.3: Item loadings for the three factors.

Item	Factor 1	Factor 2	Factor 3
active	0.80	-0.02	-0.21
aggressive	-0.02	0.71	-0.02
arrogant	0.17	0.74	-0.11
athletic	0.81	-0.01	-0.18
attractive	0.56	-0.08	0.24
boring	-0.03	0.70	-0.09
busy	0.49	0.33	0.08
cocky	0.20	0.66	0.05
committed	0.76	0.10	0.08
competitive	0.71	0.22	0.17
confident	0.63	-0.00	0.26
conscientious	0.47	0.07	0.36
consistent	0.68	0.15	0.18
controlled	0.63	0.22	-0.00
dedicated	0.78	0.11	0.03
determined	0.78	0.06	0.08
disciplined	0.78	0.13	-0.05
drinks_alcohol	-0.27	0.19	0.41
driven	0.77	0.09	0.12
eats_healthily	0.81	-0.03	-0.05
energetic	0.82	-0.02	0.02
enthusiastic	0.43	0.09	0.48
exercises_regularly	0.87	-0.01	-0.17
extroverted	0.34	0.29	0.37
fit	0.86	-0.10	-0.08
focussed	0.80	0.08	0.15
friendly	-0.12	-0.10	0.74
fun	0.00	-0.08	0.77
funny	-0.16	0.07	0.70
goal_directed	0.83	0.10	0.16
greedy	-0.33	0.64	0.10
happy	0.28	-0.07	0.54
health_conscious	0.82	-0.02	-0.10
healthy	0.83	-0.08	-0.11
inactive	-0.71	0.27	0.16
introvert	-0.29	0.31	0.25
kind	-0.19	0.01	0.66
lazy	-0.69	0.23	0.17
loud	-0.06	0.41	0.36
moody	-0.19	0.55	0.11
motivated	0.86	-0.04	0.10
obssessive	0.39	0.46	-0.03
organised	0.54	0.25	0.26
rude	-0.05	0.73	-0.04
sedentary	-0.43	0.28	0.28
serious	0.25	0.46	0.12
social	0.24	-0.04	0.55
sporty	0.84	-0.00	-0.15
strict	0.54	0.38	-0.01
strong	0.66	-0.18	0.19
unhappy	-0.39	0.48	0.05
unhealthy	-0.80	0.19	0.13
vain	0.36	0.54	-0.14

be expected, since fear of being judged and fear of being deemed incompetent are common barriers to group-based gym exercise (Ross & Melzer, 2016). There is a fine line between self-confidence and hubris. As Tracy & Prehn (2012) report, in the absence of contextual information about expended effort, pride is more likely to be interpreted as hubristic than authentic. Participants in this study were only provided with a description of a body, and no such contextual information. This resulted in somewhat ambivalent word evaluations, with confidence and dedication, and arrogance being rated in a similar direction in response to 'athletic' individuals.

Sociability The third factor included items related to friendliness and sociability, such as *Friendly, fun, funny, drinks alcohol.* This suggests that a dimension of the stereotype space includes warmth and friendliness, which occupies a different area of the warmth spectrum.

Item Selection

Factor loadings were extracted from all three sets of data. Items with a stronger correlation than ±0.45 were retained. Items that loaded strongly onto factors in the full dataset, and the two filtered datasets were retained. The items were then checked for content validity, ensuring that a range of words from the construct were present, and duplicates were avoided to reduce violations of local independence (i.e. variation should be explained at the latent variable level, not between observed variables) (Goretzko et al., 2019). Next, synonyms were removed, and in cases with antonyms , negative loaded items were included. The final set of words included 29 items, representing a variable-to-factor ratio of 9.66. KMO tests were run on full and filtered datasets using only the final 29 words. The values ranged between 'Middling' and 'Marvellous'. For the full set, sampling adequacy was calculated as KMO = 0.90, for the athletic dataset KMO = 0.77, and for the plus sized dataset KMO = 0.74.

Bayesian Ordinal Regression

To determine whether ratings of the factor words differed depending on the prime, the datasets were modelled using multi-level ordinal Bayesian regression using the `brms` package in R (Bürkner, 2017, n.d.; Bürkner & Vuorre, 2018). The factors were represented as latent variables constructed from the individual items:

$$Y_{ij} = \beta_0 + \beta_{physique} * \beta_{gender} + \epsilon \tag{6.1}$$

Ordinal regression was used because the data were categorical and applying metric models to this level of data can result in vastly inflated error rates (Liddell & Kruschke, 2018). Rather than modelling the outcome measure directly as draws from a continuous distribution, ordinal regression is a generalised linear model that allows the outcome to be modelled as the odds that a response will shift from a lower to a higher category as a function of a linear set of predictors.

Bayesian regression requires priors to be set on all parameters. Since this is a new area, weakly informative priors were placed on the intercept and beta coefficients. To test the sensitivity of the data to the priors, the models were repeated with the default brms priors, priors with a wide standard deviation ($N(0, 10)$), and priors that were tailored to the data after it was known ($N(-0.4, 0.1)$ and ($N(0.4, 0.1)$)[4]. The final priors were taken from the `auto_prior()` function from the `sjStats` package in R (Lüdecke, 2018) which returned $\beta_0 \sim N(0, 10)$ for the intercept, and $\beta_i \sim N(0, 5)$ for the slopes. Factors 1 and 3 were relatively insensitive to the priors, however the data for factor 2 was weaker, and the overall model was more sensitive to the prior.

Three models were run using Bayesian Ordinal Regression: an *intercept*$-$

[4] I refer to these priors as 'hacked' because I have no a priori reason for selecting them other than that they provided the result that I expect. With repeated testing, it is possible that the posteriors of future studies will converge on these values, but until then they are just presented as a curiosity.

only model, an *intercept+physique* model, and an *intercept+physique×gender* model. Random effects for participant, and item were modelled into the equations, giving the full equation of:

$$y \sim \alpha + \beta_{physique} \times \beta_{gender} + (1|subject) + (1|item) \tag{6.2}$$

Trace plots and posterior distributions for each parameter may be found in the appendices. All models converged, and $rhat = 1$ for all parameters. To test model adequacy, Bayes factors using Bridge Sampling (Gronau & Singmann, 2017) and Leave One Out Cross Validation (LOOCV) (Vehtari, Gabry, Yao, & Gelman, 2018) were used.

The updated posteriors from this study are to be applied as priors in future stereotype rating studies.

LOOCV

Table 6.4 presents the Leave-One-Out information criterion comparisons for each factor and model. In all cases, one of the predictor models were favoured over the intercept, however the differences were small, and the standard errors were large for each of the criterion, suggesting that more data is required before strong inferences may be made. For factor 1 and 2, the full model (physique and gender) was a better fit for the data than physique alone. For factor 3, the full model was a weaker model, suggesting that adding gender to the model as a predictor provides no new information and may over-fit the data. A frequentist interpretation would reject these values as 'significant' because they do not exclude 0 from the 95% confidence intervals. However, the results provide interesting indices with which to investigate these stereotypes further. It is anticipated that with more data, these differences will become more reliable.

Table 6.4: Cross validation model comparison between the different models.

	LOOIC	SE
F1 ICP-PHY	18.78	10.36
F1 ICP-FUL	19.12	10.37
F1 P-FUL	0.33	1.39
F2 ICP-PHY	1.99	2.68
F2 ICP-FUL	3.92	3.08
F2 P-FUL	1.93	1.74
F3 ICP-PHY	3.80	4.42
F3 ICP-FUL	3.11	4.67
F3 P-FUL	-0.69	1.59

Note. Indicates the difference in LOOIC estimates between two models. For example, ICP-PHY indicates that the physique model estimate is taken from the intercept only model. ICP = Intercept, PHY = Physique Only, FUL = Full model. A positive difference means that the first model is larger than the second, a negative difference means the opposite.

Sensitivity Check

To check the sensitivity of the likelihood to the prior distributions, the models were run multiple times with different priors. The first were wide, weakly informative priors placed on the beta coefficients, constraining the distribution to ± 10. The second was narrower, constraining 95% of the distribution between -0.8 and 0 for factors 1 and 2, and 0 and 0.8 for factor 3. These priors were selected after the data were seen and the direction of the effect was known (hence the different signs for each factor). Factor 2 was the most sensitive to the prior distribution, with a credible interval of [-1.92 – -1.38] under the wide prior and -1.51 – -1.09 under the narrow prior. Although the distributions of factors 1 and 3 also altered with narrower priors, the overall inference (i.e., direction of the effect) did not change.

Table 6.5: Regression table of the three factors with weakly informative priors

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.98	-3.25 – -2.74	-1.43	-1.81 – -1.06	-1.45	-1.83 – -1.08
3	Intercept.2.	-2.37	-2.62 – -2.13	-0.75	-1.13 – -0.38	-1.06	-1.43 – -0.69
4	Intercept.3.	-1.67	-1.92 – -1.43	-0.04	-0.42 – 0.33	-0.53	-0.89 – -0.16
5	Intercept.4.	-0.78	-1.03 – -0.55	1.03	0.66 – 1.41	0.63	0.27 – 1.00
6	Intercept.5.	-0.05	-0.29 – 0.18	1.82	1.44 – 2.20	1.51	1.14 – 1.90
7	Intercept.6.	0.79	0.54 – 1.02	2.59	2.19 – 2.99	2.52	2.12 – 2.91
8	p	-1.65	-1.92 – -1.38	-0.23	-0.66 – 0.21	0.51	0.16 – 0.86
9	g	-0.02	-0.30 – 0.26	0.43	-0.01 – 0.87	-0.28	-0.63 – 0.08
10	p.g	-0.21	-0.60 – 0.19	-0.57	-1.21 – 0.06	0.25	-0.25 – 0.73
11	Random Effects						
12	var	0.37	0.92	0.76			
13	precis	2.73	1.60	1.64			
14	ICC	0.12	0.37	0.32			
15	N	14 idx	8 idx	8 idx			
16		194 id	194 id	194 id			

Note. Scores have been converted from cumulative frequencies to implied standardised mean differences between latent distributions. The six intercepts represent the estimated positions of the k-1 cut-points on a standard normal distribution. The beta coefficients are estimated change in the latent distribution, in z scores, per unit change of the predictor.

Table 6.6: Regression table of the three factors under narrow and directed priors

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.85	-3.08 – -2.62	-1.63	-1.95 – -1.31	-1.37	-1.72 – -1.02
3	Intercept.2.	-2.24	-2.46 – -2.02	-0.95	-1.26 – -0.65	-0.98	-1.32 – -0.63
4	Intercept.3.	-1.54	-1.76 – -1.32	-0.24	-0.55 – 0.07	-0.44	-0.79 – -0.10
5	Intercept.4.	-0.66	-0.87 – -0.45	0.84	0.52 – 1.15	0.71	0.36 – 1.05
6	Intercept.5.	0.07	-0.15 – 0.28	1.62	1.30 – 1.94	1.59	1.24 – 1.94
7	Intercept.6.	0.90	0.68 – 1.12	2.39	2.06 – 2.74	2.59	2.23 – 2.96
8	p	-1.30	-1.51 – -1.09	-0.38	-0.63 – -0.13	0.49	0.26 – 0.72
9	g	-0.01	-0.21 – 0.19	0.03	-0.23 – 0.29	-0.08	-0.31 – 0.15
10	p.g	-0.43	-0.68 – -0.17	-0.29	-0.59 – 0.01	0.22	-0.06 – 0.50
11	Random Effects						
12	var	0.57	0.95	0.79			
13	precis	2.52	1.57	1.61			
14	ICC	0.18	0.38	0.33			
15	N	14 idx	8 idx	8 idx			
16		194 id	194 id	194 id			

Note. As discussed in the main text, there was no a priori reason to assume these priors. The table is presented to demonstrate the sensitivity of the data to P(rior)-Hacking.

Regression

Although each of the $K - 1$[5] cut-points has its own intercept, they each share common beta coefficients. That is, the slopes apply to each intercept. This is because cumulative frequency is modelled, and the odds across the intercepts are not independent (if 1 is less likely, then 2-7 must be more likely)[6]. Table 6.5 presents the regression parameters from the Bayesian model with weakly informative priors.

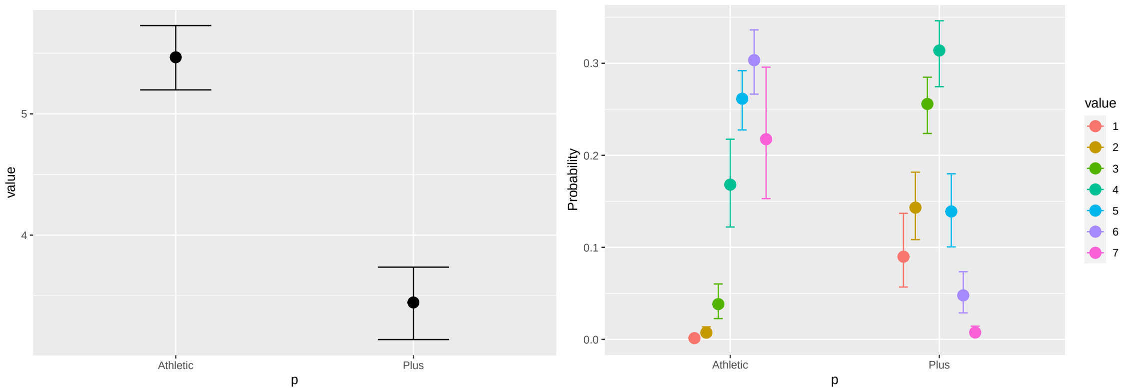
Physique

For Factor 1, there the applicability ratings of the words related to health and competence were the most affected by the physique described in the stimulus. There was a decrease in the applicability of these words of -1.65, 95%CI = [-1.92 – -1.38] on the latent scale for plus sized people, compared with people with athletic body types. See Figure 6.5a and 6.5b for visualisations. The Y axis represents the approximate cut point on the latent distribution, meaning that athletic individuals are more likely to be rated as a five on the Likert scale, whereas plus sized people are more likely to be rated as between $3.5 - 4$ (Figure 6.5a).

For Factor 2 (Arrogance and Vanity), there was a decrease in the applicability ratings of the words of -0.23, 95%CI = [-0.66 – 0.21] on the latent scale for the plus-sized group, compared with the athletic group. This suggests that the probability of a word from this factor being rated as more applicable was reduced for groups with larger bodies. However, the credible interval on this score was wide and did not exclude zero. The majority of the density is assigned to negative values, and the distribution is centred on a weak-to-moderate effect size, so there is a higher probability that the effect is negative, but these results

[5]Where K is the number of scale points, and K-1 because the cumulative probability is constrained to equal 1, and so one degree of freedom is lost.

[6]It is possible to model independent slopes for the intercept, but there was no conceptual reason that this would be necessary.

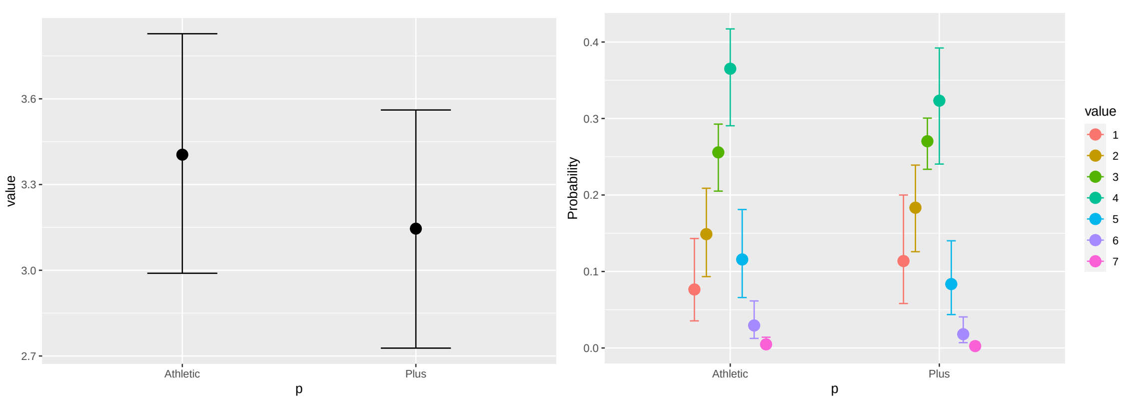


(a) Marginal effects plots for the effect of exemplar physique on Factor 1 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 1 word rating task.

Figure 6.5: Marginal plots for the effect of exemplar physique on Factor 1 applicability ratings

must be interpreted cautiously subject to further analysis[7]. See Figure **??** a & b for visualisations. There is more uncertainty surrounding the estimates, with words being more likely to be rated as applicable in athletic individuals.



(a) Marginal effects plots for the effect of exemplar physique on Factor 2 word ratings.

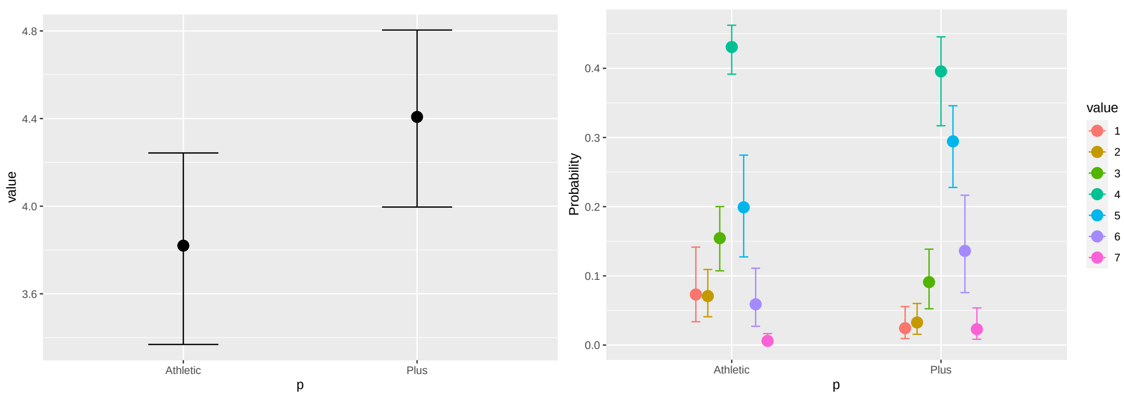
(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 2 word rating task.

Figure 6.6: Marginal plots for the effect of exemplar physique on Factor 2 applicability ratings.

For Factor 3 (Friendliness and Sociability), there was an increase in

[7]To double check, more data is currently being collected from a different population (USA summer school students), and data collection is underway for a Confirmatory Factor Analysis study using only the 29 selected word. It is unlikely that these replications will feature in the thesis, although an appendix may be added with preliminary findings.

the applicability ratings of these words on the latent scale of 0.51, 95%CI = [0.16 – 0.86], suggesting that the probability of a word from this factor being rated as more applicable increased with larger body types, compared with athletic body types. Again, there is a wide credible interval around this value, meaning that there is a large degree of uncertainty about the 'actual' value, although there is a high probability that the value is positive. See Figures 6.7a & 6.7b for visualisations.



(a) marginal effects plots for the effect of exemplar physique on Factor 3 word ratings.

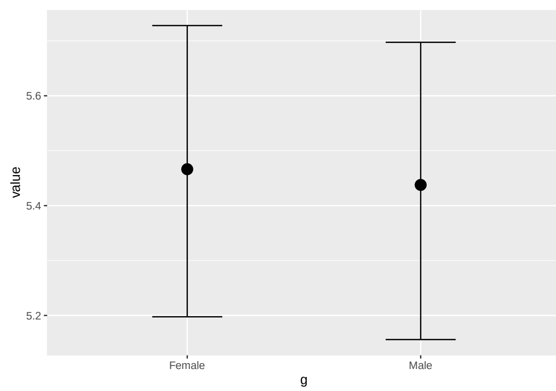
(b) marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 3 word rating task.

Figure 6.7: Marginal plots for the effect of exemplar physique on Factor 3 applicability ratings.

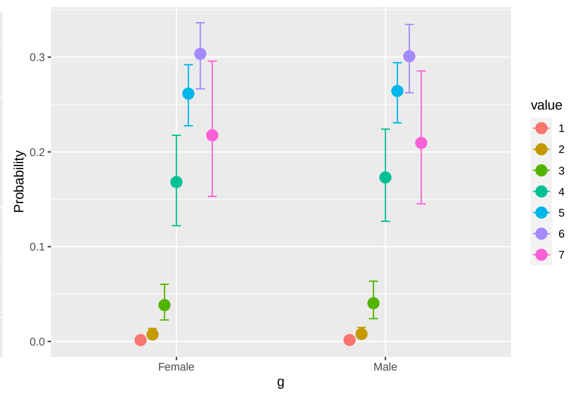
Gender

For Factor 1, there was a small effect of gender, β = -0.02, 95%CI = [-0.30 – 0.26] with a wide credible interval. This suggests that gender had a negligible effect on the applicability ratings of health competence-related words. The credible interval is wide, with the majority of probability density being assigned to values around 0. See Figure 6.8a and 6.8b for visualisations.

For the Factor 2 words, there was an increase in the applicability ratings of 0.43, 95%CI = [-0.01 – 0.87] when the descriptor stimulus was male. Again, there is a wide 95% credible interval which crosses 0, however the majority of the probability density is assigned to values above zero, suggesting that men are more



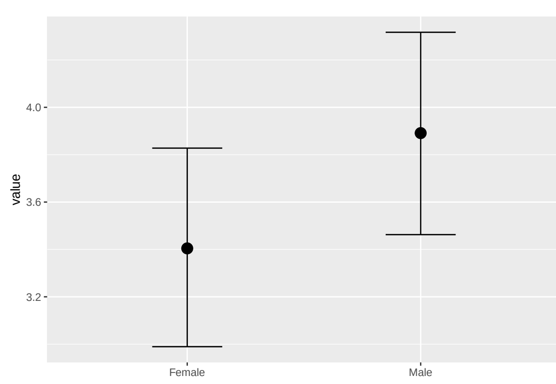
(a) Marginal effects plots for the effect of exemplar physique on Factor 2 word ratings.



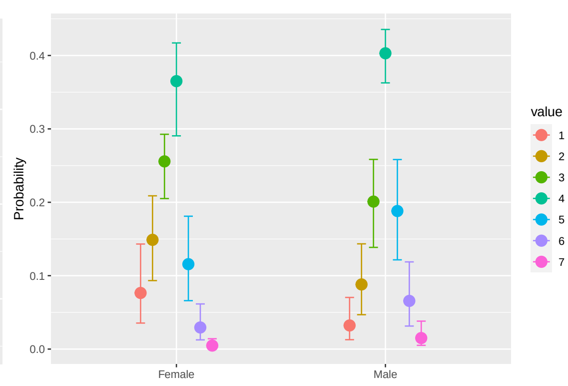
(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 2 word rating task.

Figure 6.8: Marginal plots for the effect of exemplar gender on Factor 1 applicability ratings.

likely to be rated higher for applicability than women. See Figure 6.9a and 6.9b for visualisations.



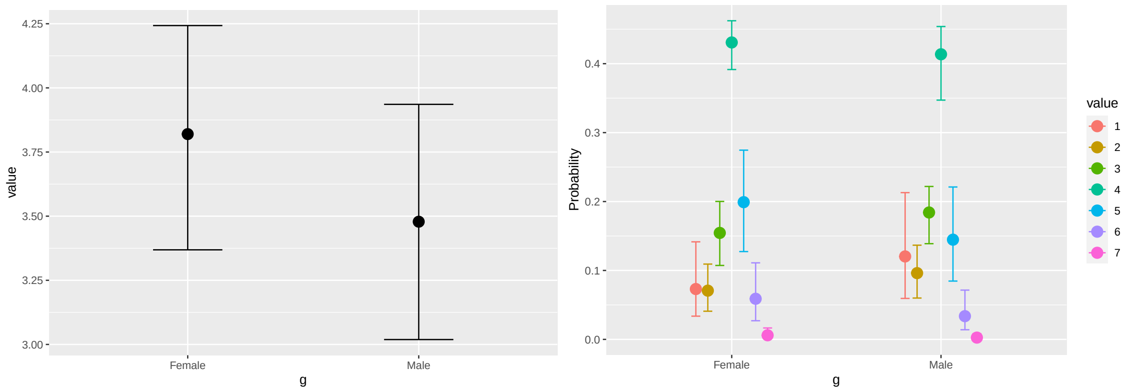
(a) marginal effects plots for the effect of exemplar gender on Factor 2 word ratings.



(b) marginal probabilities for each of the 7 Likert points for the effect of exemplar gender on Factor 2 word rating task.

Figure 6.9: Marginal plots for the effect of exemplar gender on Factor 2 applicability ratings

For Factor 3, there was a decrease in applicability ratings on the latent scale of -0.28 95%CI = [-0.63 – 0.08] when the target was male. Once again, the credible interval for this distribution was wide and included zero, although there was a greater probability that men would be rated as less friendly than women. See Figure 6.10a and 6.10b for visulations.



(a) marginal effects plots for the effect of exemplar gender on Factor 3 word ratings.

(b) marginal probabilities for each of the 7 Likert points for the effect of exemplar gender on Factor 3 word rating task.

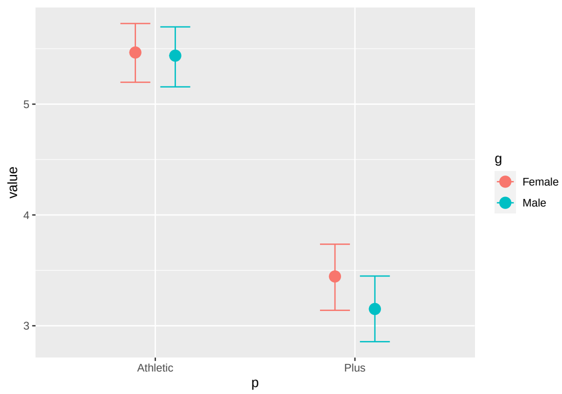
Figure 6.10: Marginal plots for the effect of exemplar gender on Factor 1 applicability ratings

Interactions

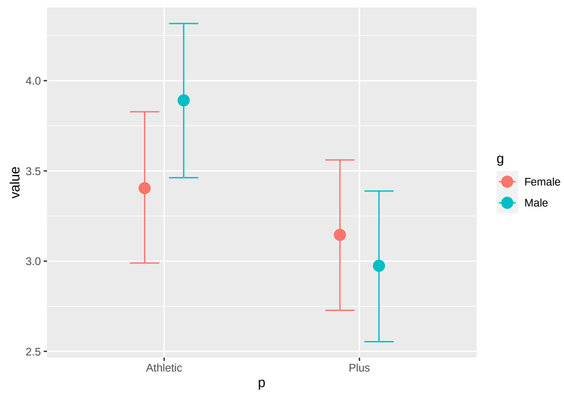
An interaction term was included in the regression equation to explore how word ratings changed with gender *and* physique. For Factor 1, there was a very minor interaction between physique and gender of β = -0.21, 95%CI = [-0.60 – 0.19]. This means that the applicability of the words was slightly higher when the target was female and plus sized than when the target was male and plus sized. This effect is presented in Figure 6.11a.

For Factor 2, there was reductive interaction between physique and gender of β = -0.57, 95%CI = [-1.21 – 0.06]. This is presented in Figures 6.11a and 6.11b. Athletic women were rated as less arrogant than athletic men, whereas the ratings of larger-bodied women was slightly higher than larger bodied men.

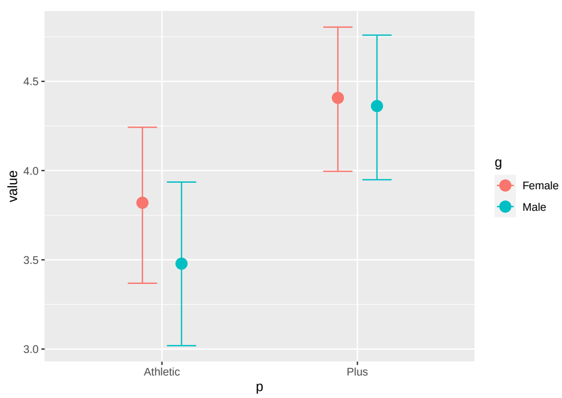
Finally, for Factor 3, there was an effect of 0.25, 95%CI = [-0.25 – 0.73] on the applicability ratings of the words related to friendliness when the model was male and plus sized. This is demonstrated in Figure 6.11c. Plus-sized men and women were rated approximately equally in terms of sociability, whereas athletic women were rated as more sociable than athletic men.



(a) Plot showing the interaction between exemplar gender and physique on Factor 1 word ratings.



(b) Plot showing the interaction between exemplar gender and physique on Factor 2 word ratings.



(c) Plot showing the interaction between exemplar gender and physique on Factor 3 word ratings.

Figure 6.11: Marginal plots for the effect of exemplar gender on Factor 1 applicability ratings

Results Summary

In this study, a preliminary, ambivalent factor structure was identified from the ratings of the applicability of words to prototypical descriptors of larger-bodied and athletic people. The structure consisted of 3 factors: Determination; Arrogance; Affability. Confidence in these factors was only reached when the full set of ratings was included in the factor analysis. These factors were then analysed using multi-level modelling to determine whether they varied as a function of the type of prime used. The most reliable effect was the effect of physique on ratings of determination words. The words that were related to determination were far more likely to be rated as more applicable to athletic groups than the larger-bodied groups. There was also a greater degree of certainty around this effect. Estimates of the between subject variability in factors 2 and 3 were less credible. There was a smaller and more uncertain effect of physique on arrogance, with athletic groups being rated as more arrogant than the plus sized groups. A stronger effect of gender was observed on these words, although zero was included in a tail of these observations. Men were rated as more arrogant than women, in general, and there was a relatively strong interaction between gender and physique, with athletic men being rated as the most arrogant, and plus-sized women being rated as less arrogant. Finally, plus-sized people were rated as being more affable than the athletic people. There was a smaller effect of gender, with men being rated as less friendly than women.

6.5 Discussion

This initial exploratory study has identified a potential stereotype structure that is related to athletic and plus-sized body-types. In broad terms, the stereotypes appear to be consistent with the Warmth-Competence dimensions described by (Fiske, Cuddy, & Glick, 2007). However, the competency factor was formed of a mixture of general and domain specific items; that is, in addition to typical items like ‘confident’, ‘determined’, items like ‘eating healthily’ and ‘exercising regularly’ also loaded on to the competence dimension. In terms of warmth, this spread over two factors. The first included words related to hubris and disrespect for others, such as: ‘arrogant’, ‘aggressive’ and ‘rude’. In the second warmth factor (Factor 3), words were warmer and related to affability and kindness (e.g., ‘fun’, ‘friendly’, ‘happy’). At this early stage, strong claims cannot be made about the nature of these stereotypes as they pertain to representations of different body types. There was some evidence that these dimensions have discriminant validity over different body types, as suggested by the beta-coefficients of the regression model. However, for Factors 2 & 3 there were wide credible intervals, indicating a high degree of uncertainty in the estimates. In Factor 3, the lower tail of the 95% CI was close to 0, and for Factor 2 it crossed zero into the potentially positive values. This suggests that more observations are required to update the current distributions. If the effect is simply noise then the centre of the distributions ought to shift towards zero, whereas if the effect reflects some level of truth then the credible interval ought to shrink, or the distribution shift away from zero. That is to say this research line is still in its infancy, and further research is needed to explore the magnitude of the effect.

One of the purposes of this exploratory study was to update the weak prior belief for future studies into athletic stereotypes. In the models described above, only very weakly regularising priors were used. This allowed the MCMC algorithm to explore the wide space between ±10 for the slope, and ±20 for the intercept parameters (the points between which 95% of the probability density lies), assigning only 5% of the prior probability to values $\alpha_{ki} \geq 20| \leq -20$ or $\beta_i \geq 10| \leq -10$[8]. The majority of these values are incredibly unlikely in the current context. A Cohen’s d of 5 or greater is almost unheard of in social science. The largest meta-analytic effect size from 100 years of social psychology as reported by Richard et al. (2003) was r = 0.75, which is equivalent to a mean difference of d = 3.43. As more data is collected, the prior distribution will start to reflect a more

[8]That is, we were confident that the beta would not be of a magnitude of ±10-∞

reasonable parameter space. The intention is to use the posterior distribution from the current study as an informed prior in future studies, including exploring the activation of the stereotypes in response to visual representations of the groups.

Warmth-Competence

Interestingly, the stereotypes identified through Factor analysis were broadly consistent with the universal stereotype content model described by Fiske, Cuddy, Glick, & Xu (2002). This model consists of two dimensions: warmth and competence. When evaluating people based on group membership such as ethnicity or occupation, word generation and evaluations often group around constructs related to friendliness and kindness, and intelligence and efficacy. According to Fiske et al. (2007), these dimensions are orthogonal: that is, they will correlate positively for some groups (e.g., Christians, middle classes, homeless people; Fiske et al., 2002), and negatively for others (the elderly: high warmth, low competence; the British; low warmth, high competence; Cuddy, Fiske, & Glick, 2007). There is some evidence to suggest that this model is culturally invariant within 10 non-US cultures; with the exception that there may be less in-group favouritism within collectivist cultures (people from an Asian country were less likely to rate their own group as positive on both dimensions; Cuddy et al., 2009). The orthogonality of these dimensions means that ambivalent (mixtures of positive and negative traits) stereotypes are entirely possible. Indeed, words rated in response to the athletic condition were evaluated as more confident and dedicated than in response to the larger bodied condition, but also more arrogant, vain, and less friendly.

Hubris, Pride, and Envy

A possible antecedent to the factor structure is that the primes evoked feelings of envy. Envy would require a desire to look athletic, and in the absence of individual measures of athletic internalisation (Schaefer et al., 2015; Thompson & Stice, 2001), this is too big a leap at this time, and is a point for further investigation.

Factor 2 included words that were related to hubris, such as arrogance, vanity, and rudeness. Since there were higher ratings of this factor in response to athletic body types, it would be interesting to explore whether exemplars of these groups evoke the 'conjugal' feeling of envy, as predicted by Dual Envy theory (Crusius & Lange, 2014; Lange & Crusius, 2015; Van de Ven, Zeelenberg, & Pieters, 2012). Under this model, envy is separated into malign and benign types. The envy response can be predicted by the 'type' of pride that a model displays, that is pride and envy interact, with displays of pride from the 'superior other' (the envied person) determining to some extent the reaction of the perceiver (Lange & Crusius, 2015). Pride is a feeling of joy in response to a status and may be affected through countenance and posture. Pride may also be separated into two categories: hubristic, and authentic. Hubristic pride refers to pride in the global self, rather than of individual events, and may be perceived as unearned. Authentic pride is related to pride in a specific event, when the status has been earned rather than assigned (Tracy & Robins, 2007). In Lange & Crusius' *Tango of Two Deadly Sins*, these two types of pride may evoke different types of envy. The display of hubristic pride evokes feelings of malign envy, in which attention is drawn towards a model, and the equalisation of statuses is associated with bringing down the target. Malign envy is associated with feelings of Schadenfreude[9], and with desires to 'level the playing field' by doing harm to the 'other' (Lange, Weidman, & Crusius, 2018). Compared with the experience of benign envy, the experience of malign envy draws attention away from the task that a model is performing (or has performed), and towards the individual themselves (Tracy & Robins, 2007). In contrast, authentic pride invokes benign envy, in which attention is drawn to the effort exerted to succeed at a task. Focus on the task is associated with motivation to succeed, 'levelling the playing-field' through effort

[9]German word for the enjoyment in instances of the suffering of 'superior' others.

and self enhancement. There is a suggestion that the extent to which the status is evaluated as deserved influences the type of envy (Van de Ven et al., 2012), with malign envy being experienced if the target is deemed to not deserve their status. From a health-promotion perspective, interventions that elicit malign envy could be harmful, since it is associated with a shift of attention towards the model rather than towards the goal (Crusius & Lange, 2014).

Perception of hubristic pride in advertising models may affect perception of the item or service and have an anti-advertising effect. For instance, Sung & Phau (2019) discuss perceived hubristic vs authentic pride as eliciting malicious vs benign envy when participants observed luxury advertising. That is, if a model is seen as arrogant, then malicious envy is experienced by the audience and attitudes towards the brand are reduced. Making the jump from advertising and social dynamics to health psychology, there may be consequences of including models that seem hubristic in health-promotional materials. If athletic body-types induce malign envy, then the audience's attention may be drawn away from the benefits of the task, and towards the various ways that the model could be made less-athletic.

Weight Stigma

There are both notable differences and similarities between the current work and previous work on weight stigma and fat phobia. The Fat-Phobia scale was developed between 1984 and 1993 and similar dimensions related to dedication and commitment were found, however larger-bodied people were deemed to be unfriendly and unsociable (Robinson et al., 1993). The Fat-Phobia scale has since been reduced to a single dimension consisting of 14 items which exclusively focus on the dedication and commitment attributes of people with larger bodies and seems to be valid when compared with the full version (Bacon et al., 2001). However, the revisited the Fat-Phobia scale did not include the ideographic component used in the original development; that is, the scale represents current ratings of words generated in the 1980's. The most common words that were generated, and subsequently rated in the present study, frame larger bodied people as more friendly and social than athletic individuals. This may represent a shift in social perceptions of larger bodied people or could just mean that athletic people are just seen as less friendly. In the former case, it could be that the prevalence of the ‘friendly fat person’ trope in popular media has led to a shift in perceptions, however the current dataset has insufficient detail to explore this possibility.

Strengths and Limitations

The thorough development of the visual stimuli that were evaluated was a strength of the current study - although these will not be used until Chapter 7. The study also benefitted from a factorial design, in which participants were blinded to the purpose of the study. Finally, the adherence to recent suggestions means that the ordinal data generated during this study were handled in the (current) least error-prone way.

A limitation of this study is that it was not feasible to analyse whether word ratings were affected by gender matches or mismatches. It is feasible that applicability ratings of participants own gender would differ, although there is no literature (to the best of The Author’s knowledge) on how these ratings would differ, and to what extent. In a future study, words will be rated in response to exemplars of both genders. Sample size was a further issue in this study, since it was too small to make strong inferences. The results should really only be regarded as a training dataset to make future predictions. The group was largely made up of undergraduate students from the United Kingdom and Norway, as well as responders from Twitter, although there is no way to determine the proportions, since demographics such as occupation and ethnicity were not collected. What is known is that, as noted in section 6.3, the sample was heavily biased towards young women. It is debatable whether knowledge of stereotypes (meta-stereotypes) can be expected to vary across genders, and there has been no investigation of this to date. As more data is accrued on this topic, it will be feasible to explore this possibility.

Applied Context

A consequence of this study for the applied context is that the assumptions that are widely made about the evaluation of athletic individuals require updating. They may not be ubiquitously positive, a belief that was presumably formed the basis of many health advertising campaigns. Traditionally it has been assumed that athletic individuals are revered, with no discussion of negative associations that there may be (see page 47). This may be applied to promotional materials, since the presentation of athletic individuals in gym advertisements are intended to elicit benign envy but may actually result in malign envy if the model is evaluated as arrogant or hubristic (page 222). This concept has been demonstrated in luxury product advertisements (Sung & Phau, 2019), but not promotional materials for health-related behaviours (this point will be expanded further from page 259). As of yet, only prototypical evaluations have been addressed, and in Chapter 7 and Chapter 8 whether these stereotypes are also associated with visual stimuli will be explored.

Future Directions

Future studies will identify whether there is a similar stereotype structure if participants are presented with visual stimuli (i.e., the exemplars produced in the pre-study). The contribution of gender to the ratings will also be studied. Men are traditionally seen as naturally stronger than women, and so a greater degree of effort may be attributed to women becoming athletic, than men (Messner, 1988). Once the stereotype structure identified in this study has been confirmed in the visual modality, attempts will be made to alter ratings by providing additional information. Further, the contribution of envy and perceived pride could also be explored in future studies, although this falls outside of the scope of the current thesis.

Findings and Conclusions

Previously, stereotypes of people with athletic bodies (as discussed in Chapter 2, page 47), were largely based on conjecture, and little research has been conducted on evaluations of this group. Stereotypes of people with larger bodies were mostly negative. Although model fit was unsatisfactory when the data were divided into subsets, the potential stereotype structure that has been identified for athletic bodies and plus sized bodies in this study suggests a different *out there* stereotype structure.

The structure, formed of the three factors (health and competence, arrogance, and affability), was broadly consistent with the warmth-competence dimensions suggested by (Fiske et al., 2002). The stereotypes also have positive and negative elements and appear to be a mixture of positive and negative elements, which runs counter (or at least adds complexity to) to the quote from Ginis & Bassett (2012) on page 47 stating that those who exercise are revered. Moreover, the *dumb jock* stereotype discussed in Chapter 2 (page 47) did not emerge as a salient stereotype for the more general group of *people with athletic bodies.*

Relation of Data to Proteus Effect

Words related to competence and health, including dedication and conscientiousness, were rated as more applicable to athletic people than plus-sized people, but so were words associated with arrogance and hubris. The factor structure was broadly constant between physiques, but with ratings of words clustering towards opposite poles of the Likert scale. These findings bode well for the research on the Proteus Effect and exercise-related behaviours discussed in Chapter 4, since the necessary conditions of the Proteus Effect (e.g. the existence and evocation of salient stereotypes) can potentially be met (see page 52). Thus far the stereotype structure has only been supported in response to prototypical descriptions, and more work is needed to determine whether the structure is evoked by visual representations.

Conclusion

This study identified salient stereotypes associated with athletic and larger bodies. A dichotomy was broadly supported, with athletic people being rated higher than larger bodied people on the Determination and Arrogance dimensions, but lower on Affability dimension. This supports the previous assumptions made by researchers investigating the Proteus Effect on exercise, with the caveat that the structure is more complex, and if the Proteus Effect is robust, then negative attitudes might be assimilated into the player's temporary self-schema too. Where support for a $Laziness|Physique$ activation is supported by the new structure, other activations may also occur, such as $Arrogance|Physique$ or $Friendliness|Physique$, not to mention the various blended activations that may also occur (e.g. what would a $Laziness \times Arrogance|Physique$ activation look like?). This chapter only investigated the prototypical representations, and there is still work to be done translating these stereotypes into the visual representations required by the Proteus Effect. In the upcoming chapter, the words will be re-rated for applicability by a new group of participants in response to the visual representations described in section 6.3.

Chapter 7

Looking Lean and Seeming Mean: Stereotypes and virtual exemplars of athletic and plus-sized individuals

7.1 Rationale

In Chapter 6, a tentative stereotype structure was identified for male and female athletic and plus-sized *prototypes*. The study was based predominantly on responses from women from a younger age group. Given the ambivalent stereotype space that these prototypes occupy, inferences about how exemplars of these prototypes may influence behaviours may start to be drawn. Prior to this, however, a visual stimulus that sufficiently represents the prototype must be validated. In the current study, the words identified and refined in Chapter refwordex were rated against visual instances (synthetic humanoid models) of the 'athletic' and 'plus-sized' prototypes. This represents the next step towards creating valid avatars to use in a Proteus Effect study.

7.2 Introduction

In the phenomenon known as the Proteus Effect, the appearance of a user's avatar has been shown to temporarily alter self-perception, and in-turn affect their behaviour (Yee et al., 2009). However, the effect should only occur if the appearance of the avatar contains salient information from which behaviour may be inferred. The size of an avatar's body has been used as a manipulation to alter the amount of effort expended by users whilst playing an exergame (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014). The assumption is that when avatars are perceived as lazy, a common stereotype related to larger bodies (Puhl & Heuer, 2010; R. Puhl & Suh, 2015; Robertson & Vohora, 2008), the player's self-perception is temporarily shifted, and they believe that they are lazy too, leading to less exertion during the exergame. The three studies that investigated this *laziness|physique* activation described in Chapter 4 solely relied on the in-game avatar customisation options of existing exergames. Li et al. (2014) used the Nintendo Wii game EA Sports Active to create 'normal'[1] and 'obese' avatars for their participants; and Peña et al. (2016) and Peña & Kim (2014) used the game 'Virtua-Tennis' to create 'normal' or 'obese' avatars.

As discussed in Chapter 4, these avatars were not validated for this purpose. The avatars in the Peña et al studies were designed using only Body Mass Index (BMI) as a reference; and in the Li et al study there was a brief mention of a 'pre-test', but no information on what this entailed was included. A further limitation is that these studies only considered the consequences of using larger-bodied avatars on in-game performance rather than assessing a spectrum of body sizes or types. Following the conditions addressed on page 52, the Li and Peña studies made the following assumptions: i. stereotypes are available for the group that the avatar represents; ii. the stereotypes associated with the appearance of the avatar are univalent (i.e. only positive, or only negative); and iii. the avatars they designed both sufficiently represented and carried enough

[1]Their words.

information to portray the stereotype. However, Chapter 6 suggests that both athletic and plus-sized stereotypes are *ambivalent*.

7.3 Ambivalent Stereotypes

Stereotypes are often more complex than the Proteus Effect research would imply. Indeed, there is a body of work on *ambivalent stereotypes*, in which evaluations may be independently positive or negative on different dimensions (Fiske et al., 2002). This means that there may be positive stereotypes associated with larger-bodied individuals, and negative stereotypes associated with athletic bodied-individuals. This is tentatively supported by the findings in Chapter 6, in which participants evaluated the group of individuals with athletic body types as competent, but arrogant and unfriendly; and people with plus sized body types as friendly but incompetent.

The Stereotype Content Model (Cuddy et al., 2009; Fiske et al., 2002) suggests that stereotype structures may differ at group and individual levels. It states that there are two universal dimensions of stereotypes: warmth and competence; and that these correlate negatively at a group level. However, because of a bias towards the consistency of evaluations (i.e. the so-called 'halo effect'), they may correlate modestly positively when evaluations are made towards an individual (Fiske et al., 2007). The current study aims to determine whether a similar ambivalent structure of body-related stereotypes is evident when participants evaluate synthetic individuals (exemplars). However, there are a range of complexities when forming directional hypotheses on this topic.

Depictions of Larger Bodies

There is a mixture of findings surrounding the impact of counter-stereotypical information on evaluations of larger bodied people. In a study that used an *Imaginary Contact* paradigm in which participants were asked to imagine larger bodied people with positive attributes, negative weight bias was reduced (Dunaev, Brochu, & Markey, 2018). However, in a study on visual portrayals of 'obesity', 'obese' and 'non-obese' participants were presented with images of larger bodied and athletic people in stereotypical, neutral, and counter-stereotypical contexts. Non-obese individuals reported higher levels of 'fat-phobia' when judging images of larger bodied individuals both exercising and performing stereotype-congruent behaviours such as sitting and eating junk food (Pearl et al., 2015a), i.e. that these representations were equivalent for 'non-obese' groups[2] (Pearl et al., 2015a). They also reported that there were no differences in fat phobia when only 'obese' individuals were analysed[3]. The authors also found that, while the non-overweight participants both liked and felt more comfortable exercising than 'obese' people, observing images did not significantly affect this measure in either group[4].

The above study provides some limited evidence about the effect of context on evaluations of larger-bodied people. In the current study, Figures were presented in isolation, with no environmental or behavioural context. The virtual exemplars were standing neutrally, wearing sports-wear[6]. In addition, male and female models were used, whereas in the Pearl et al study only women were depicted.

Pearl et al. (2015a) reported no significant effects of model race on evaluations, however this is not to say that race differences should not be explored further, but in this study, only white models were used. On a methodological note, it is important to consider that the data the Pearl et al study were analysed at an interval level, when an ordinal scale may potentially be more appropriate. As discussed in Chapter 3 (page 82), it has been found that the use of metric methods on ordinal constructs can vastly inflate the false

[2]Although no test for equivalence was conducted.

[3]Despite the authors reporting non-significant tests as findings.

[4]The authors report uncorrected pairwise comparisons made after a non-significant interaction. This inflates the false positive (type-1) error rate, and even if it were significant, the p value of 0.03 would not have survived the 16-fold Bonferroni correction. [5]

[5]The authors state that *no other pair-wise comparisons were significant*

[6]of sorts.

positive and false negative (type-1 error rate, type-2) error rate, or entirely reverse the known effect (Liddell & Kruschke, 2018).

The Current Study

The current study sought to validate the stereotype-evoking properties of the set of exemplars created in Chapter 6 using the word rating task from Chapter 6. These exemplars represented plus-sized and athletic body types and are to be used in future studies[7].

The following hypotheses for the confirmatory phase of this study were pre-registered on the Open Science Framework prior to data collection.

- *H1* The factors identified in the exploratory phase (Chapter 6) will be replicated (as indicated by adequate model fit) using Confirmatory Factor Analysis when rated by a new group of participants. These factors will be Determination, Arrogance, and Affability.
- *H2* Participants responses to the athletic and plus sized exemplars will be negatively correlated such that athletic exemplars will be rated as more Competent and Arrogant and less Affable than plus-sized exemplars.
- *H3* Ratings of words associated with the factors will show discriminant validity between participants responding to plus sized and athletic body types as shown using Bayesian Ordinal regression, with priors taken from the posterior distribution of the exploratory analysis from Chapter 6. Words related to arrogance and determination will be rated higher in response to athletic bodies, while words related to Affability will be rated higher in response to plus-sized bodies.

[7] In each case, avatars were white. This decision was made not to imply a 'default' skin colour, but to avoid racial intersectional stereotypes associated with skin colour, as discussed in previous research (e.g. Sailes (1993)). Moving forward, the racial sensitivity of these body types will be addressed, since the FaceGen Artist software can provide models and textures representing a range of races.

7.4 Method

The method and analysis plan for this study were pre-registered on the Open Science Framework. The registration and materials can be found at https://osf.io/hnpax.

Participants

Inclusion Criteria

To be eligible for this study on Prolific, participants were required to be over the age of 18, have normal to corrected-normal eyesight, and be able to understand written English. Since this study was concerned with meta-stereotypes, the inclusion criteria was kept broad, and no specific knowledge or experience was required.

Sample Size

A sample size of 300 respondees on the crowed sourcing platform Prolific.ac was chosen for this study, and these were recruited from Prolific.ac. This number was imposed due to limited resources available. Due to an error with the survey, the first 100 participants responded to words for three of the conditions on a 6 point, rather than a 7 point scale meaning that 75 had to be excluded. The analysis was conducted on the 225 participants who responded to the first exposure on a 7 point scale. There were sufficient samples to run the Bayesian analysis, however the model fit estimates for the full CFA model and EFA phase were poor, so a further 102 participants were recruited from Prolific and the analyses were re-run leaving a final sample size of 327. Participants were reimbursed £1.00 for completing the study[8]. In total, 327 participants were included in the analysis. Note again that the sample size was below the suggested minimum of 400 (Goretzko et al., 2019).

[8]Earning the equivalent of £8.32 per hour of participation

Table 7.1: Age and gender of the participants in this study.

	Female	Male	Non-binary
18 - 24	35	68	0
25 - 34	54	69	2
35 - 44	16	24	0
45 - 54	24	10	0
55 - 64	10	8	0
65 - 74	2	2	0

Demographics

Age and gender frequency data are presented in Table 7.1. This sample was slightly more balanced than the previous quantitative chapters, although in this case there was a male majority, and the distribution was once again skewed towards younger age groups. The differing demographics between the studies may be a source of bias, but again it is unclear whether meta-stereotypical beliefs vary across genders.

Apparatus

The avatars described in Chapter 6 (page 189) were used as stimuli. There were two male and two female avatars, one of each had a plus sized body, and one of each had an athletic body. The avatars were presented in a survey, with images resized so that they would fit on the screen with the word for evaluation and the scale. The survey was presented using the Qualtrics survey platform (Snow & Mann, 2013). The images were resized to fit on the screen with the target words below 7.1.

The survey was also optimised to run on smartphones, tablets, and computers.

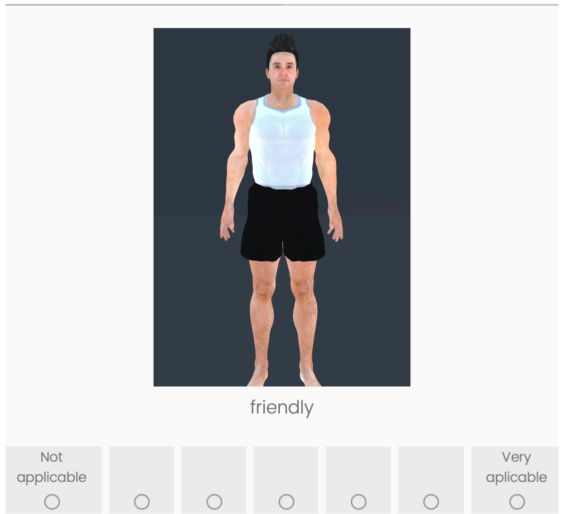


Figure 7.1: Example of athletic male exemplar stimulus

Procedure

After reading the information sheet and providing their informed consent, participants were informed that they would be rating words in response to a presented image. The same 53 words from Chapter 5 were presented sequentially in a random order for participants to evaluate on a 1-7 scale with only two anchor points (1 - not applicable, 7 = very applicable). To reduce the number of clicks required for each trial, questions progressed automatically after each response. The back-button was disabled in all trials. In total, there were 4 awareness checks included in the study. Participants were asked to respond “Very applicable” in these cases. Inclusion in the analysis was reliant upon consistently passing these checks, however the maximum number of misses across all participants was 1, and so no cases were omitted. Due to a mistake when making the survey, 75 participants responded to stimuli on a 6 point scale. These participants were excluded from the analysis. Each participant responded to two exposures; one athletic and one plus-sized. Both exposures were of the same gender. The order of each was randomised, and balanced. In total, there were 480 observations.

Analysis

In the initial CFA, and first EFA section of the analysis, only the first ratings were used. The confirmatory phase of the analysis involved Confirmatory Factor Analysis (CFA) to determine whether the factor structure identified using the 29 items identified in Chapter 6 replicated with visual stimuli. The same ordinal regression analysis was also run on the single observations from each participant. Next, to test the hypothesised within-participant ratings differences between physiques, ordinal regression was run on the within-subject IV, with gender of the avatar as the between-subject IV. The exploratory phase of the analysis involved searching for alternative factor structures within the visual-stimuli dataset. This was repeated for first exposure and second exposures, prior to being combined into a large dataset. Finally, the first exposure data from the visual condition were combined with the verbal descriptor data and regressions were run between the two types of stimuli for each factor. This was done to determine whether evaluations differed between the two types of stimuli.

7.5 Results

The confirmatory results section follows a pre-registered design which can be found at osf.io (https://osf.io/hnpax).

Sampling Adequacy

First exposure

Sampling adequacy for the first exposure dataset was determined using Kaiser's Little Jiffy Mk4 (AKA Kaiser, Meyer, Olkin index). The total KMO scores were 0.97, for the athletic dataset KMO = 0.92, and for the plus sized dataset KMO = 0.92.

Second exposure

Sampling adequacy for the Second exposure dataset yielded a KMO for the total set (both physiques) of 0.97, for the athletic dataset KMO = 0.88, and for the plus sized dataset KMO = 0.82.

Both exposures The scores values for both exposures participants yielded a KMO of 0.98, for the athletic dataset KMO = 0.94, and for the plus sized dataset KMO = 0.94.

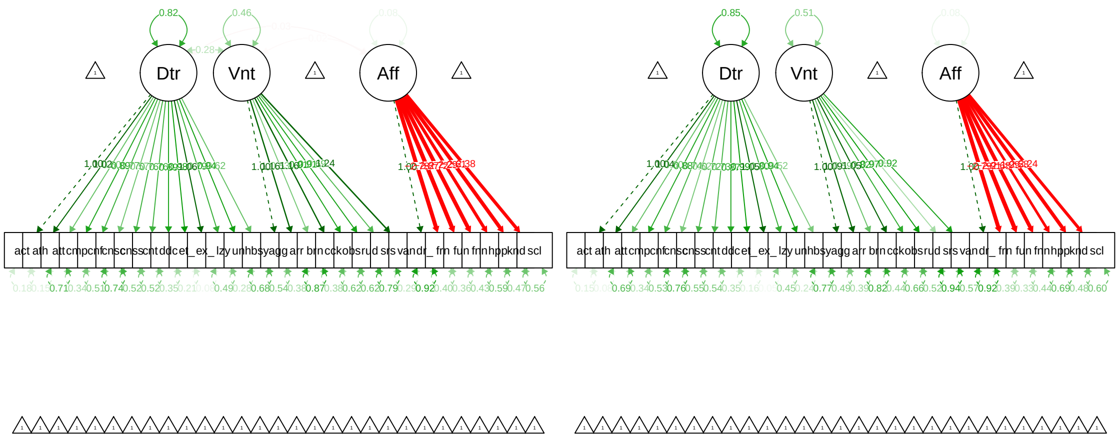
CFA

Confirmatory factor analysis was run on the set of 29 words Chapter 6. The model fit was not adequate for the single (first) observations (RMSEA = 0.095). It is likely that there was an insufficient sample size for the full model, which had 60 free parameters to estimate. Instead, the full model was decomposed into three separate factors, meaning that there were fewer overall parameters to estimate. The fit indices were good for each factor (note that this was not the case for the smaller sample size, in which factor three was above the threshold of 0.08). The factor diagrams are presented in Figure 7.2a and 7.2b[9].

Regression

As in Chapter 6, ordinal probit regression was run on the visual data to determine whether the factors from the verbal phase also varied depending on the physique and gender of the exemplar. In addition, the same models were run with both default and weakly informative priors from Chapter 6 as a sensitivity check.

[9]It is important to note however that the decomposed factor model was not pre-registered and should be interpreted as exploratory. Further, the pre-registered sample size was exceeded, the analysis was run twice, and the results of the first analysis informed the continued collection of data. Although optional stopping has not been discussed in the literature in terms of CFA (yet; to the best of The Author's Knowledge.), it is possible that the error rates have been inflated due to this practice.



(a) Full CFA model, RMSEA = 0.095

(b) Individual CFA models for each factor. They are presented on the same panel for brevity but were not entered into the analysis together

Figure 7.2: CFA Models.

Sensitivity Check

To check the sensitivity of the likelihood to the prior distributions, the models were run multiple times with different priors. The first priors were taken from the posterior distribution from the verbal phase of the study (Chapter 6). Using the `fitdistrplus` package in R (Delignette-Muller & Dutang, 2015), the mean and standard deviations from each slope parameter ($\beta_p, \beta_g, \beta_p : g$) were estimated and included as priors in the new brms model (see table 7.4). The second set of priors were the wide, weakly informative priors from Chapter 6 were placed on the beta coefficients, constraining the distribution to ± 10.

The posterior estimates were affected by the prior distributions, but the inferences remained the same. There was a relatively strong reduction in Determination ratings when the avatar had a larger body. Factors 2 and 3 were relatively sensitive to the prior, but the direction of the posterior slope did not change with the prior. Factor 2 had a credible interval of -0.47 – 0.04 under the weak prior and -0.53 – -0.14 under the informed prior. Although the distributions of Factors 1 and 3 also altered with narrower priors, the overall inference (i.e. direction) did not change.

As in Chapter 6, Factor 2 was particularly affected by the choice of prior. Under the weakly informative prior, there was a difference in slope parameter of 0.18, compared with the informed prior. However, in this case, zero was not included in the 95% credible interval. For Factor 1 there was also a difference in slopes between the weakly informative and informative priors, this was a large difference (b = 0.23), however the confidence in the estimate was also improved using the informative prior. The beta estimate for Factor 3 was relatively invariant under both prior conditions, although confidence in the estimate was improved slightly in the informed priors.

Table 7.2: Regression table of the posterior parameter estimates for the effect of exemplar physique and gender on the three factors, under weakly informative priors

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.37	-2.61 – -2.14	-1.55	-1.88 – -1.22	-1.69	-1.94 – -1.45
3	Intercept.2.	-1.62	-1.86 – -1.40	-0.81	-1.13 – -0.48	-1.08	-1.32 – -0.85
4	Intercept.3.	-1.03	-1.26 – -0.81	-0.20	-0.52 – 0.13	-0.51	-0.74 – -0.28
5	Intercept.4.	-0.19	-0.42 – 0.03	0.92	0.60 – 1.26	0.73	0.50 – 0.97
6	Intercept.5.	0.53	0.30 – 0.75	1.74	1.42 – 2.09	1.66	1.42 – 1.90
7	Intercept.6.	1.27	1.05 – 1.50	2.50	2.16 – 2.85	2.35	2.10 – 2.60
8	p	-1.44	-1.64 – -1.24	-0.22	-0.47 – 0.04	0.20	-0.02 – 0.42
9	g	0.62	0.43 – 0.81	0.51	0.27 – 0.76	0.37	0.15 – 0.59
10	p.g	-0.74	-1.01 – -0.47	-0.48	-0.84 – -0.14	-0.50	-0.81 – -0.19
11	Random Effects						
12	var	0.33	0.63	0.47			
13	precis	3.15	1.59	1.32			
14	ICC	0.10	0.28	0.26			
15	N	14 idx	8 idx	7 idx			
16		327 id	327 id	327 id			
17	Marginal R2 / Conditional R2	0.417 / 0.553	0.061 / 0.374	0.016 / 0.263			

Note.

Table 7.3: Leave One Out Cross Validation Information Criterion (LOOCVI) comparison between the three models

	LOOIC	SE
F1 ICP-PHY	43.24	12.99
F1 ICP-FUL	49.50	14.04
F1 P-FUL	6.26	3.36
F2 ICP-PHY	4.04	4.82
F2 ICP-FUL	6.26	5.54
F2 P-FUL	2.21	2.41
F3 ICP-PHY	-0.10	0.67
F3 ICP-FUL	1.05	2.15
F3 P-FUL	1.15	2.19

Model Selection

For each factor, Leave One Out Cross Validation, LOOCV, (Vehtari et al., 2018) was applied to the three models (intercept, physique, all factors). The differences can be found in table 7.3. In all cases, the full model offered the better fit[10]. It is important to note that the differences in LOO IC were not particularly large for Factor 2 and 3. In only Factor 3 did adding gender contribute substantially to the fit of the model. The full models for each factor will be used in the following results section.

Model Descriptions

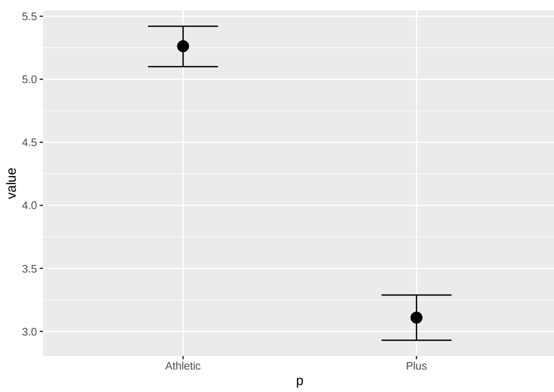
Table 7.4 provides the parameter estimates for each factor in the full model (physique and gender).

[10] only in Factor 1 did this exceed the frequentist criteria for 'significance' (i.e. $2 \times StandardError$)

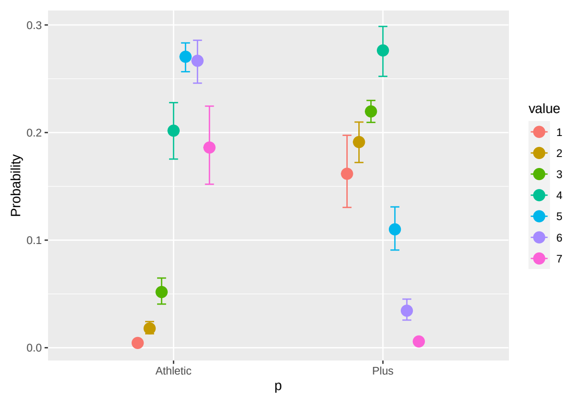
Table 7.4: Regression table of the posterior parameter estimates for the effect of exemplar physique and gender on the three factors using the posterior distributions from the parameter estimates in Chapter 6 as priors

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.52	-2.73 – -2.30	-1.67	-2.00 – -1.35	-1.76	-1.98 – -1.54
3	Intercept.2.	-1.77	-1.98 – -1.56	-0.93	-1.25 – -0.61	-1.16	-1.37 – -0.94
4	Intercept.3.	-1.18	-1.39 – -0.97	-0.32	-0.64 – -0.00	-0.58	-0.79 – -0.37
5	Intercept.4.	-0.34	-0.54 – -0.13	0.80	0.48 – 1.12	0.65	0.45 – 0.87
6	Intercept.5.	0.38	0.17 – 0.59	1.62	1.30 – 1.95	1.58	1.37 – 1.80
7	Intercept.6.	1.13	0.92 – 1.34	2.37	2.04 – 2.71	2.27	2.05 – 2.51
8	p	-1.62	-1.77 – -1.47	-0.34	-0.53 – -0.14	0.14	-0.04 – 0.31
9	g	0.37	0.22 – 0.52	0.29	0.10 – 0.47	0.18	0.00 – 0.35
10	p.g	-0.42	-0.62 – -0.21	-0.25	-0.50 – -0.01	-0.26	-0.50 – -0.02
11	Random Effects						
12	var	0.35	0.65	0.48			
13	precis	3.13	1.55	1.31			
14	ICC	0.10	0.30	0.27			
15	N	14 idx	8 idx	7 idx			
16		327 id	327 id	327 id			
17	Marginal R2 / Conditional R2	0.410 / 0.552	0.045 / 0.372	0.005 / 0.261			

Physique Under the full model, applicability ratings of the Factor 1 words were the most affected by the physique of the exemplar (see Figure 7.3a and 7.3b). People rated *Determination* word applicability ratings β = -1.62 standard deviations lower on the hypothetical latent distribution in response to the larger bodied exemplar. The 95% credible interval for this parameter was 95%CI = -1.77 – -1.47, meaning that, given the prior information and the current data, there is a 95% probability that this physique has a large negative effect on evaluations of competence-related words.



(a) Marginal effects plots for the effect of exemplar physique on Factor 1 word ratings.

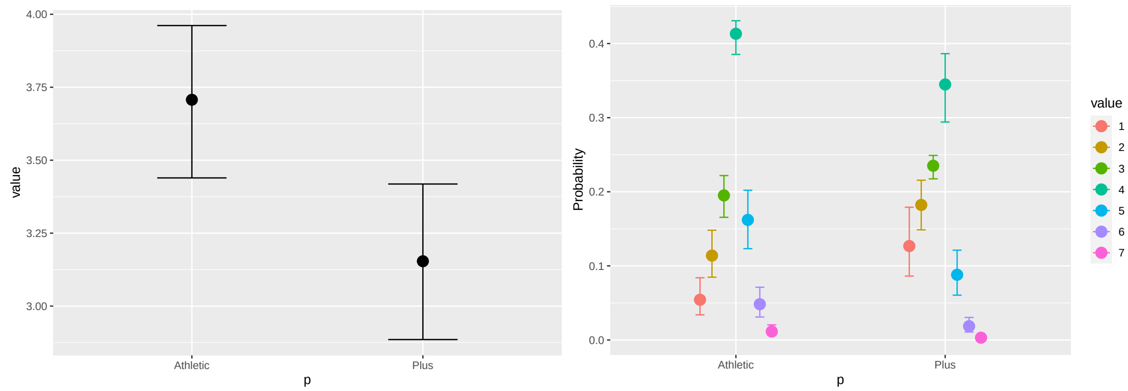


(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 1 word rating task.

Figure 7.3: Marginal effect plots for the effect of exemplar physique on Factor 1 word ratings.

There was also a decrease in applicability ratings for words related to the arrogance factor (Factor 2) in the larger bodied condition. Physique was associated with a reduction of β = -0.34 standard deviation units on the latent scale. The credible interval was 95%CI = -0.53 – -0.14, meaning that we can be very confident that the effect of plus-sized bodies on these evaluations is negative and moderately sized (see Figures 7.4a & 7.4b).

There was an increase in applicability evaluations of words related to *Affability* when the exemplar had a larger body. There was an increase of β = 0.14 standard deviation units in the applicability evaluations of these words. The lower tail of the 95% CI was very close to zero, meaning that an effect of

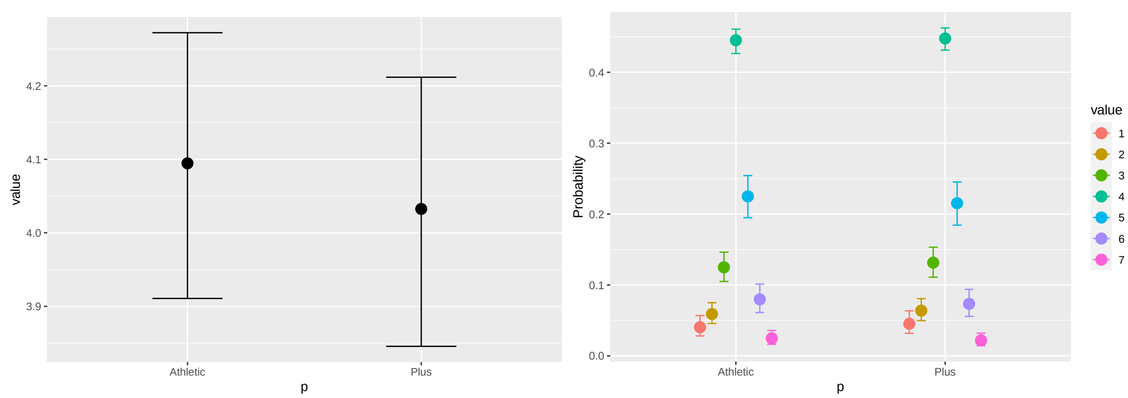


(a) Marginal effects plots for the effect of exemplar physique on Factor 2 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 2 word rating task.

Figure 7.4: Marginal effect plots for the effect of exemplar physique on Factor 2 word ratings

zero is a plausible value, however the majority of the probability density (i.e. the mean of the posterior distribution) suggests that a moderate effect is more likely. The 95%CI was -0.04 – 0.31 (see Figures 7.5a and 7.5b).



(a) Marginal effects plots for the effect of exemplar physique on Factor 3 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 3 word rating task.

Figure 7.5: Marginal effect plots for the effect of exemplar physique on Factor 3 word ratings

The direction and strength of the effects of exemplar physique found in Chapter 6 were broadly replicated in the visual avatar condition (although a direct comparison will be made in the exploratory section). Words related to Arrogance, Affability, and Determination were rated higher for men than women,

Table 7.5: RMSEA, SRMS, and TLI for the full model

	RMSEA	lower	upper	confidence	RMSR	TLI
Full data set	0.08	0.08	0.09	0.90	0.04	0.87

but the 95% credible interval for Factor 3 included 0 was no evidence of an effect of gender on ratings of friendliness-related words.

There were interactions with similar strengths between physique and gender in all three factors. Words related to Determination and Arrogance were rated lower for plus sized men, than for plus sized women. There was no evidence for an interaction for ratings for Affability-related words.

EFA

As a sanity check, EFA using parallel analysis was run on the visual dataset to ensure that the factor solution identified in Chapter 6 actually replicated. This was conducted on both subsets of exposures 1^{st} and 2^{nd}.

The factors from Chapter 6 were broadly replicated, which is surprising given that the CFA model had convergence issues and a solution could not be found.

Next, parallel analysis was run on the second exposure to determine whether there was within-subject variability in the factor loadings.

Both Modalities

A final set of models looked at how the modality of the stimuli interacted with the physique of the exemplar in the production of evaluations. The priors for these models were the same weakly informative priors used in Chapter 6 (i.e. $\beta_i \sim N(0,5), \alpha_i \sim N(0,10)$). This is because there was no prior information available on the differences between visual and verbal exemplars, to the authors knowledge,

Table 7.6: Item loadings for the three factors for the second exposure.

Item	Factor 1	Factor 2	Factor 3
active	0.94	-0.07	-0.06
aggressive	0.25	0.59	-0.15
arrogant	0.40	0.61	-0.13
athletic	0.95	-0.05	-0.09
attractive	0.74	-0.07	0.15
boring	-0.20	0.63	-0.08
busy	0.64	0.18	0.20
cocky	0.52	0.49	-0.06
committed	0.79	0.09	0.05
competitive	0.83	0.14	0.07
confident	0.80	0.10	0.06
conscientious	0.51	0.14	0.30
consistent	0.61	0.19	0.15
controlled	0.78	0.11	0.07
dedicated	0.81	0.11	0.07
determined	0.80	0.15	0.07
disciplined	0.86	0.06	0.06
drinks_alcohol	-0.52	0.29	0.29
driven	0.75	0.21	0.02
eats_healthily	0.94	-0.04	-0.08
energetic	0.89	0.03	0.05
enthusiastic	0.58	0.04	0.36
exercises_regularly	0.96	-0.01	-0.08
extroverted	0.47	0.21	0.19
fit	0.94	0.00	-0.09
focussed	0.80	0.15	0.09
friendly	-0.05	-0.15	0.73
fun	0.16	-0.11	0.66
funny	-0.05	-0.04	0.60
goal_directed	0.81	0.10	0.10
greedy	-0.49	0.32	0.15
happy	0.50	-0.01	0.36
health_conscious	0.94	-0.04	-0.01
healthy	0.94	-0.15	-0.01
inactive	-0.90	0.17	0.17
introvert	-0.39	0.28	0.24
kind	-0.16	-0.12	0.68
lazy	-0.82	0.24	0.05
loud	0.13	0.43	0.14
moody	-0.18	0.47	0.17
motivated	0.86	0.09	0.06
obsessive	0.32	0.39	0.04
organised	0.66	0.19	0.24
rude	0.06	0.64	-0.17
sedentary	-0.68	0.19	0.25
serious	0.34	0.33	0.14
social	0.54	-0.05	0.39
sporty	0.93	-0.05	-0.07
strict	0.70	0.23	0.00
strong	0.76	0.07	-0.04
unhappy	-0.54	0.34	0.08
unhealthy	-0.90	0.15	0.09
vain	0.60	0.32	-0.12

Table 7.7: Item loadings for the three factors for the first exposure.

Item	Factor 1	Factor 2	Factor 3
active	0.91	-0.01	-0.03
aggressive	0.18	0.62	-0.07
arrogant	0.23	0.55	-0.09
athletic	0.90	0.05	-0.09
attractive	0.51	-0.01	0.19
boring	-0.19	0.55	-0.13
busy	0.45	0.05	0.25
cocky	0.22	0.61	-0.07
committed	0.72	0.06	0.08
competitive	0.74	0.13	0.13
confident	0.62	0.15	0.20
conscientious	0.42	0.03	0.24
consistent	0.62	0.08	0.23
controlled	0.64	0.03	0.01
dedicated	0.78	0.03	0.08
determined	0.79	-0.01	0.11
disciplined	0.83	0.03	0.05
drinks_alcohol	-0.55	0.27	0.26
driven	0.66	0.15	0.12
eats_healthily	0.87	-0.02	-0.04
energetic	0.83	0.05	0.04
enthusiastic	0.51	0.11	0.35
exercises_regularly	0.91	0.09	-0.10
extroverted	0.27	0.16	0.36
fit	0.89	0.00	-0.05
focussed	0.72	0.04	0.09
friendly	-0.00	-0.19	0.67
fun	-0.04	-0.04	0.65
funny	-0.14	0.10	0.61
goal_directed	0.78	0.07	0.13
greedy	-0.47	0.44	0.12
happy	0.33	-0.01	0.47
health_conscious	0.83	0.05	-0.03
healthy	0.88	-0.06	-0.08
inactive	-0.82	0.10	0.11
introvert	-0.30	0.14	0.13
kind	0.01	-0.08	0.58
lazy	-0.75	0.16	0.14
loud	-0.15	0.44	0.27
moody	-0.07	0.47	0.10
motivated	0.82	-0.03	0.14
obsessive	0.20	0.47	-0.00
organised	0.55	0.09	0.23
rude	-0.07	0.57	-0.04
sedentary	-0.73	0.12	0.23
serious	0.23	0.20	-0.09
social	0.22	0.07	0.52
sporty	0.89	0.02	-0.02
strict	0.52	0.25	0.04
strong	0.69	0.08	0.04
unhappy	-0.49	0.28	-0.07
unhealthy	-0.87	0.10	0.09
vain	0.41	0.42	-0.01

and so the HMC algorithm was given the freedom to explore a wider parameter space. The posterior distribution from this study may be used to inform future investigations on this subject.

Table 7.8: Regression table of physique by experiment model. The priors were weakly informative.

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.92	-3.13 – -2.72	-1.63	-1.94 – -1.34	-1.68	-1.88 – -1.48
3	Intercept.2.	-2.23	-2.43 – -2.03	-0.92	-1.22 – -0.63	-1.20	-1.39 – -1.00
4	Intercept.3.	-1.60	-1.80 – -1.40	-0.28	-0.58 – 0.01	-0.66	-0.85 – -0.46
5	Intercept.4.	-0.75	-0.95 – -0.55	0.82	0.53 – 1.12	0.50	0.30 – 0.69
6	Intercept.5.	-0.03	-0.23 – 0.17	1.63	1.33 – 1.92	1.35	1.15 – 1.55
7	Intercept.6.	0.75	0.55 – 0.95	2.38	2.07 – 2.68	2.07	1.87 – 2.28
8	p	-1.69	-1.88 – -1.50	-0.50	-0.76 – -0.24	0.33	0.14 – 0.54
9	exp	-0.20	-0.37 – -0.03	0.19	-0.03 – 0.42	0.02	-0.15 – 0.20
10	p.exp	-0.18	-0.42 – 0.06	-0.00	-0.32 – 0.33	-0.40	-0.67 – -0.15
11	Random Effects						
12	var	0.37	0.76	0.48			
13	precis	2.99	1.57	1.52			
14	ICC	0.11	0.33	0.24			
15	N	14 idx	8 idx	7 idx			
16		521 id	521 id	521 id			
17	Marginal R2 / Conditional R2	NA	0.044 / 0.414	0.014 / 0.249			

The aggregated effect of physique on word ratings for the Factor 1 words (i.e. when both datasets were included in the model) was -0.18 standard deviation units on the latent scale, that is, strongly negative. The 95% confidence interval did not cross zero (-0.42 – 0.06), and there is a 95% probability that the effect is negative, given the data and the prior information. That is, across both modalities, there was a very strong negative effect of larger bodies on applicability ratings. For Factor 2 (Arrogance) the aggregated effect was 0, (-0.32 – 0.33). There was a decrease in ratings for Arrogance overall in the visual condition, For Factor 3 (Affability), there was a positive relationship between physique and applicability ratings -0.4 (-0.67 – -0.15), with the applicability being higher for larger bodies.

The effect of modality on word applicability ratings was not as strong, and the 95% CI was close to or included zero in all three factors. The applicability of competence and arrogance words was rated higher for visual exemplars than for the descriptors, but there was a high probability that the effect of modality on friendliness ratings was zero or close to zero. Finally, there was a potential an interaction between stimulus modality and physique for Factors 1 and 3. This was most robust for Factor 3, in which ratings were lower for visual-plus sized exemplars than athletic-verbal descriptors. For Factor 1, the applicability of the words was rated as lower for the visual depictions of larger bodied people than the descriptors of athletic people.

Results Summary

In this study, the factor structure from Chapter 6 was replicated, with determination, arrogance, and affability factors being weakly supported by CFA, and replicated by EFA methods. Moreover, the regression analysis from Chapter 6 was also somewhat replicated, however there were some differences. The effect of physique upon the ratings of the Determination words remained strong, however gender appeared to contribute more towards the determination ratings in the visual compared with the word stimuli, and this was accompanied by a stronger interaction between these two variables. Confidence in the effect of physique, gender, and the interaction between these two variables on Arrogance related words was increased in the visual condition. Finally, confidence in the effect of physique, gender, and the interaction between these two variables was reduced for Affability-related words. These differences were further explored by comparing both of the modalities, and support was found for the visual condition being rated as slightly less determined, and moderately less Affable than the prototypical word condition. There was no effect of modality on Arrogance ratings.

7.6 Discussion

Broadly speaking, the factor structure identified in Chapter 6 was replicated with visual stimuli. This is unsurprising, since the words were generated in response to both the verbal and the visual representations. However, it is useful to confirm that not only do the words map onto similar latent structures in response to both types of stimulus, but the variation remains constant across modalities.

The discussion will cover both the confirmatory and exploratory results.

Confirmatory

CFA

This was ambiguous, and the pre-registered criteria of RMSEA = 0.08 was not met in the full CFA model. It is likely that this is because there were too few data points. The KMO sampling adequacies were ‘magnificent’, and when the pre-registered 29 words were entered into Exploratory Factor Analysis, the factor structure from Chapter 6 was replicated. When the number of parameters was reduced by decomposing the full factor structure into its constituent parts (i.e. removing covariance estimates between latent variables), the model fit was good in each case. This may be appropriate in the current case, because each factor was treated as independent in the regression model but will pose a problem should multivariate analysis be employed.

Effect of Physique and Gender

The direction and strength of effects found in Chapter 6 were similar to those found in the current Chapter. Using the prior information from Chapter 6, parameter estimates for the slopes had narrow 95% credible intervals that did not cross zero, meaning that we can conclude that there is a high probability, give the information provided, that the true parameter range is non-zero. The signs and degree of uncertainty surrounding the parameter estimates remained relatively robust under the provision of less informed (and therefore more uncertain) prior distributions, but particularly for Factor 2 (the likelihood of which was also weak in Chapter 6), there was variation in the posterior estimates as a function of the prior distribution. In future studies investigating these words, sensitivity analyses will continue to be employed, and the existing posterior will be updated so that their posterior estimates continue to converge upon a stable estimate of the true effect distribution.

Exemplars with athletic physiques were judged to be more competent, more arrogant, and less friendly than those with larger body types. These ambivalent stereotypes, primed by visual information, have not been discussed in the body image or social psychology literature to the best of The Author's knowledge. This stereotype structure has important implications for depictions of health and fitness, as well as broadly supporting previous findings reporting negative behavioural responses to 'superior' others in health and fitness settings, as well providing an additional theoretical framework to explain these reactions. For instance, a potential explanation of the effect found by Wasilenko et al. (2007) is that the perception of competence, judgement/arrogance, and lack of friendliness in ‘athletic’ others may interact, increasing exercise anxiety, and the fear of judgement reported as one of the barriers to exercise in both Chapter 5 and by Pridgeon & Grogan (2012). Elsewhere in the social psychology literature, displays of hubristic pride (akin to arrogance) have been related to the inference that the ‘superior’ status was not earned through hard work, but given through luck or circumstance. This has been associated with the motivation to bring the superior other down (malign envy), rather than to work hard to attain a similar status (Crusius & Lange, 2014; Lange & Crusius, 2015; Lange et al., 2018; Sung & Phau, 2019). By this rationale, presenting an exemplar as less ‘naturally’ gifted will make the goals seem attainable, promoting benign envy, and the motivation to improve one’s situation (rather than destroy the others). Extending further, presenting exemplars as friendly may mean that the activity (exergame, gym membership, etc) is perceived as a more inviting and less judgemental environment. This is clearly a blue-skies derivation, especially since motivation and envy were neither measured in Chapter 6, Chapter 7, nor will they be measured in the future Chapters, since this is beyond the scope of the thesis. However, it would be an interesting future direction, particularly given the responses towards the athletic avatars in Chapter 5 (e.g. a smug smirk irritated one participant; whereas another saw the avatar’s body as a goal to work towards).

Interestingly the gender effects observed in Chapter 6 were weaker in the current study, and in some cases the opposite in Chapter 7. Further, the model comparison approach did not favour the full *physique* × *gender* model for Factor 3. It is possible that the visual representation of the athletic woman did not match the internal representation that participants constructed, since athletic women are generally portrayed as more slender in the media, rather than muscular. Indeed, in a content analysis of social media posts tags Talbot, Gavin, Van Steen, & Morey (2017) found that 81% of female bodies under the #fitspiration tag had ‘thin’ body types, whereas 46% had muscular body types.

Due to the weakness of gender in predicting change, in Chapter 8, the within-subject factor will be gender - i.e. participants will rate both male and female exemplars). This will allow a closer look at whether a person's gender identity affects ratings in applicability ratings of the different genders.

Exploratory

Exploratory analyses included checking the extent to which evaluations varied between stimulus modalities. There was little evidence of an effect of the modality on evaluations for Determination and Affability factors, however the visual exemplars were rated higher on the Arrogance factor. This is interesting because all four avatars had the same facial expression, and this may have come across as 'smug', a feature that was reported as undesirable and irritating by a participant in Chapter 5. There was little evidence that modality interacted with physique for Factor 1, although the most probable values were negative, meaning that larger people in the visual condition were rated as less competent and warm across all factors. Factor 3 had the strongest interaction, which may suggest that the exemplars did not represent the participant's prototypical 'friendly fat person'. Another possible explanation is that participants adjusted their ratings to be more consistently positive or negative across the stereotype space when presented with an exemplar rather than a group descriptor. That is, when the stimulus was visual, and the physique was larger, ratings for Determination and Affability were lower. This could be interpreted as an example of the halo effect (Nisbett & Wilson, 1977), however that is a question for another research programme. It would be a useful future direction to run a multivariate analysis on the data to see how the different factors interact, however the necessary software for this is still in development.

General Discussion

The exemplars that were designed in Chapter 6 appear to elicit similar evaluations to the internal representations of 'athletic body' and 'plus-sized body' group members elicited by textual descriptions. That is, the words vary in a similar way between modalities. The representations are ambivalent, meaning there are positive and negative elements to both, and broadly vary across competence and warmth dimensions, that is the athletic exemplars were rated as more competent and less warm than the plus sized exemplars.

An Ontological Perspective

The relations between verbal and visual representations of the exemplars can be modelled in ontological terms (see page 64). Under this framework, the verbal descriptors ('People with Larger Bodies') are *Classes*, and the exemplars are *instances* of these classes. Each class has a set of attributes (the latent variables) which may be measured using the ratings of the individual words. That the factors from Chapter 6 replicated supports the assertion that the visual exemplars can be seen as instances of the class of athletic and larger bodied individuals. In addition to attributes, classes (and by extension, individuals) have relations between them. The regression analyses from Chapter 6 and Chapter 7 support the existence of relations between the classes of plus sized and athletic bodied people, and individuals of these classes. The relations were that plus sized people are rated as less competent and arrogant, but more friendly than athletic-looking people. This perspective provides a useful language for future studies. In Chapter 8, additional attributes will be added to the larger bodied exemplars (*is active*, or *is not lazy*) to determine whether the negative relations between athletic and plus sized individuals are weakened. In Chapter 9, I will explore whether these instances can directly affect behaviour in an exergame. Participants will ostensibly embody the exemplar (making it an avatar) and play an exergame. This will require that participants hold the abstract class of plus or athletic bodied people which has the required attributes, that the exemplar is a suitable instance of these classes, that the participant's self-perception is altered and they temporarily adopt the attributes of the instance (and therefore class) as their own[11].

Strengths and Limitations

The sample in this study was not constrained to psychology students and was drawn from a large and diverse pool (in terms of age and gender, no other demographics were taken). The study also benefitted from a pre-registered design, meaning that The Author was constrained to interpret the results as intended, rather than presenting *ad hoc* adjustments as confirmatory.

However, the initial sample size was suboptimal, and more participants were added based on the fit of the models. For the Bayesian elements of the study, this is less of an issue since by adding more data, confidence in the estimate is increased (Royall, 2017), however for the factor analysis sections this may pose a problem. When p values contribute towards the evidence of a study, adjustments may be made (providing these are stated *a priori*) to the alpha level to account for multiple analyses (e.g. Pocock, 1977). There is, however, a dearth of advice on error rates due to multiple analyses in model comparisons methods.

A potential limitation with the current study is the lack of validated measures of appearance-related bias, however part of the purpose of the current study was to further validate the wordlist used in Chapter
refwordex. Although future studies may assess convergent validity of the word ratings with other measures of 'fat-phobia', the constructs do broadly align. The Fat Phobia scale has 14 pairs of antonyms and assesses participants general

[11]IYI An anecdotal note. A further observation is that the instantiation process appears to not be lossless, as suggested by the weaker effects of physique in the visual condition compared with the verbal condition. This is interesting because it was also reflected in the word generation process. Participants produced a more diverse range of words in the athletic condition when responding to verbal descriptors of groups than visual exemplars. In the plus sized condition, there was little difference. No inferential analysis was conducted on this, and again it is beyond the scope of the project. However, for interested readers, word clouds from each condition are provided in the appendix https://osf.io/6mqsv/.

attitudes towards 'fat' people using two antonymic anchors and a 5 point Likert scale (Bacon et al., 2001). The measure consists of words such as *active* and *inactive* which are represented in the 53 words used in the current study.

A further limitation was that the within subject factor of physique meant that participants were un-blinded to the purpose of the study in the second exposure. As a result, there was considerable variation in the evaluations between participants with different orders of presentation. In the next study, participants will rate both genders rather than both physiques.

The prior information that the analysis from this study was based on was potentially biased, since nearly 75% of the sample in Chapter 6 were women. This means that the Bayesian updating may be inaccurate. However, the EFA and CFA did not use prior information[12] and similar structures to those found in Chapter 6 were identified independently. Moreover, the slopes of the responses were similar to those in Chapter 6, both under weak and informative priors. Again, whether there are variations between genders at the meta-stereotype level is unclear and is something that will need to be addressed in future research.

A potential limitation to the study is also the use of a crowdsourcing platform. Although steps were taken to ensure that surveys were completed by 'aware' participants (awareness checks etc), there is still a possibility that some 'bots'[13] may have mined the funds. Prolific have numerous safeguards in place to prevent such abuses (Bradley, 2018), but The Author acknowledges the possibility that some of the responses were fabricated. It is unclear how sophisticated such bots are, but it would not be trivial for such bots to correctly identify the purpose of the study in the absence of text (images and videos were used in the past two studies) and simulate the hypothesised data.

[12] In the Bayesian sense anyway. Arguably the CFA model was based on prior information.
[13] software that can complete surveys.

Applied Context

This Chapter has demonstrated that the complex stereotypes associated with the prototypes discussed in Chapter 6 also apply to visually presented, synthetic exemplars. This was foreshadowed in the comment by one of the participants in Chapter 5, that their athletic avatar looked like a 'smug jock'. The exemplars in this study had neutral facial expressions, and so the contribution of a "smug smirk" towards negative evaluations is yet to be determined.

In applying the findings to a wider context, many forms of advertising rely on benign envy to promote goods and services (Belk, 2011). Benign envy, as discussed previously (page 222), is evoked when the target is viewed as having earned their status, generally by affecting authentic pride. The use of athletic models in health-related advertising (e.g. gym adverts) are presumably selected on the basis that they will be seen to have earned their status. However, given that malign envy is elicited through the perception of hubristic pride, an athletic exemplar attributed as affecting arrogance effecting an apposite adverse envy is an appropriate assumption. Although this has not been applied directly to health and wellbeing, nor has it been applied to avatars, there is evidence in the para-social interaction literature (of which some authors have argued that avatar relations are a subset of (J. Banks & Bowman, 2016a), that malign envy may be negatively associated with ability social-comparison (the degree to which social network users compare their successes and failures with a models ability) related inspiration (Noon & Meier, 2019), the very construct that advertisers are attempting to enhance. Further, and potentially more importantly, Noon & Meirer discuss the contribution of homophily (self-similarity) of targets in promoting positive social comparisons. They demonstrate that an increased perception of self-similarity (homophily) with social networking *others* (e.g. influencers) was associated with a reduction in the malign envy that people reported experiencing as a result of ability social comparison. As mentioned earlier (Chapter 6 page 222; Chapter 7 page 253), malign envy is associated not with motivation for personal growth, but with motivation to bring down the superior other.

In relating this back to avatars, if a person experiences their avatar as *an other* rather than having an *as self* relation (page 22), then malign envy may occur if perceived homophily is low. The envy theories discussed from page 222 would predict that this would result in attempts to cause the avatar to fail, rather than for the self to succeed. Although any semblance of confirmation for this suggestion is far off, it may be relevant to the previous research discussed in Chapter 4, in which participants were assigned avatars based solely on body size, with no controls for homophily, or identification style. Indeed, as discussed in Chapter 5 , facial similarity was not sufficient to promote homophily, and when participants created *ideal others* nor was it necessary. To put it bluntly, simply superimposing a person's face on a muscular body[14] will not always lead to identification with that avatar. However, the availability of more flexible affordances in avatar creation, or a more nuanced approach to tailoring might help. For example, subtle facial similarities have been used to influence voting intentions. In a study by Bailenson, Iyengar, Yee, & Collins (2008), photographs of politicians faces were blended with photographs of participants. When presented to non-partisan participants, these blended images were associated with higher intention to vote for the 'similar' candidate. This may be something to consider for cases in which players engage with their avatars parasocially, rather than under the 'avatar as self' relation (J. Banks & Bowman, 2016a). Clearly, more work is needed to link the complex network of interactions between homophily, envy, avatar identification, and motivations, to avatars in health-related behaviour, but these possibilities should be explored prior to large scale acceptance of body-type restrictions in avatar systems.

[14]like a cat-fisher's dating profile image

Relation of Data to the Proteus Effect

As in Chapter 6, evaluations for both athletic and plus sized exemplars were ambivalent. Prior to applying this to the Proteus Effect (Chapter 9), in Chapter 8, attempts will be made to break the negative, Determination-related evaluations of the larger bodied exemplars by introducing additional diagnostic information to the characters. The rationale is that if this is possible in the stereotypical ratings of avatars, then if the Proteus Effect holds, behaviour and self-perception ought to be affected in a similar way. Using prior information from previous research on the effect of stereotype-incongruent information on evaluations, and an estimate of the magnitude of the effect of diagnostic information on evaluations, for the first time in research on the Proteus Effect it will be possible to directly determine the effect of avatar/exemplar appearance on evaluations and behaviour.

Findings and Conclusion

In addition to the activation by prototypical descriptors (Chapter 6), the activation of the three-factor stereotype by visual exemplars of larger-bodied and athletic individuals was supported by the findings of this study. Moreover, since the factor structure found in the EFA section of Chapter 6 was replicated in the current study, this is taken as support for the generalisability of the stereotype beyond exclusively undergraduate groups. The current sample was also more representative of different ages and had a more balanced male/female ratio, although was not representative of other genders. The successful evocation of the three-factor stereotype by the visual exemplars suggests that the stimuli developed in Chapter 6 meet the second necessary condition discussed in Chapter 2 (page 52), that an avatar must carry stereotypical information has been met. Finally, previous literature on stigma related to body shapes discussed in Chapter 2, page 44 suggests that larger-bodied people are only evaluated negatively and appear to have missed the ostensibly positive stereotype related to Affability that was found here and in Chapter 6. This study has demonstrated that positive, in addition to negative stereotypes may be evoked through visual stimuli depicting larger bodied people. The stereotypes structures and physique-related variations found in Chapter 6 were replicated in this chapter using the visual representations designed in section 6.3. This suggests that the exemplars contain sufficient visual information to evoke *out there* stereotypes.

Chapter 8

Speed and weight!: The effect of counter-stereotypical information on impression formation

8.1 Rationale

Chapters 6 and 7 tentatively identified an ambivalent stereotype structure both for the prototypical descriptors and synthetic instances (exemplars) of 'people with larger body types' and 'people with athletic body types'. Having demonstrated salient stereotypes, predictions of behavioural assimilation through the Proteus Effect may be made. It would be expected that gamers playing with athletic body types would perceive themselves as dedicated and active ($dedication|physique$), which would lead to greater exertion during an exergame. Conversely, people embodying larger bodied avatars would be expected to perceive themselves as 'lazy' and inactive ($lazy|physique$), resulting in less exertion during exergaming, compared with the athletic group. This much has been demonstrated in previous studies (see Chapter 4, although now there is a solid rationale for explaining the resultant behaviours, assuming the Proteus Effect was the cause. Given the critical stance taken throughout this thesis, confirming this and concluding that larger bodied avatars should indeed be eschewed would be unsatisfactory. The purpose of this chapter is to lay the foundations for a subsequent hypothesis: if health-related behaviours are affected by the stereotypical attributes of avatars, then altering these stereotypes at the impression formation stage will moderate the behavioural modification. The following sections will introduce theories, and present research supporting assertions that stereotypical evaluations may be manipulated. Following these foundations, an experiment was conducted in which the exemplars from Chapter 7 were imbued with animations, in an attempt counteract negative impression formations of larger bodied exemplar.

8.2 Introduction

[1] In Chapter 6, a group stereotype structure was posited for athletic and plus sized individuals. In Chapter 7, the structure was tested again using visual exemplars of these groups. This Chapter examines whether appearance-based evaluations of individuals may be altered by providing further diagnostic information about a group representative. In brief, negative evaluations of larger-bodied individuals, such as laziness and low self-control, will be challenged through the depiction of an exemplar[2] running on the spot showing no signs of fatigue, and an exemplar standing stationary. This study broadly draws from research on impression formation and maintenance, particularly in reference to the competence-morality/warmth dimensions of group perception. It will use a novel paradigm of providing evidence in favour of a larger bodied person being fit. According to previous research, 'negative' stereotypical beliefs about a target may be counteracted during impression formation by providing additional 'positive' information (Paolini & McIntyre, 2019). As will be discussed, researchers have suggested that the diagnosticity of positive and negative information

[1]The Chapter title is one of the Tekken 6 and 7 character Bob's victory quotes

[2]This term is used to describe the models in this study, which are conceptually distinct from avatars (user controlled) or agents (computer controlled). The latter have some degree of 'intelligence', whether through user or algorithm.

is asymmetrical. That is, negative information has a more dramatic effect on evaluations than positive information (Fiske et al., 2007; Skowronski & Carlston, 1987, 1992; Wojciszke, 2005)[3]. This suggests that provision of positive information should alter evaluations, but not as much as the provision of negative information.

Media Depictions Of Larger Bodies

Two content analyses of online media portrayals of larger-bodied people identified that members of this group are frequently presented in a stigmatised way (Heuer et al., 2011; Puhl et al., 2013). For example, larger-bodied people are often presented performing unhealthy behaviours such as eating and drinking high sugar or fat products; dressing in undersized clothing; and in unflattering poses with attention drawn to the abdomen or backside. The authors note that these models are often 'anonymised' by way of being 'headless'. This representation is similar in both photographs and videos of larger-bodied people. A later, qualitative, content analysis investigated portrayals and treatment of larger bodied individuals on the television show 'The Biggest Loser' (Greenleaf et al., 2019). The authors report stereotype endorsement, aggression, and the disregard of the contestants experiences as recurrent themes in the discourses throughout. Contestants were depicted as lazy, unattractive, and incompetent; shouted at and pushed around; and expected to adopt a 'no-pain no-gain' mentality. Puhl et al. (2013) notes that such portrayals reinforce negative stereotypes, and that without positive depictions, the stigmatisation is unlikely to change. Moreover, attribution of blame for fatness is focussed on the individual, rather than the complex network of associations between media, society, and psychology, all of which need to be addressed (Greenleaf, Chambliss, Rhea, Martin, & Morrow Jr, 2006; Greenleaf & Weiller, 2005). Using what was found in these content analyses, a set of images and videos with positive, stereotype-challenging depictions of larger-bodied people

[3] There are some conditions for this observation, which will be discussed below.

exercising was developed by the Rudd Center for Food Policy and Obesity. These images include larger bodied people shopping for vegetables, being active, and working in an office environment in 'business' clothing (Rudd-Center, 2019)[4].

Images from the Rudd Center were used in a subsequent study in which participants were presented with samples from the set and asked to complete measures of 'fatphobia' and exercise intentions (Pearl et al., 2015a). The authors found limited evidence for an effect of mere exposure to the images (see discussion in Chapter 7). However, it is unlikely that simply viewing an image is sufficient to immediately change attitudes. Indeed, Ferguson (2013) suggested that the immediate, negative effects of media exposure on self-perception or body satisfaction, are small, and limited to 'at-risk' individuals. It is therefore likely that diffuse and frequent representation of fit and healthy larger-bodied people is likely needed for any real change to occur. In the current study, the methodology used in typical media exposure studies was informed by social-cognitive theories on stereotypes and impression formation.

Theoretical Background

Cue Diagnosticity

Earlier work on cue diagnosticity framed evaluations in terms of honesty and intelligence, rather than the warmth-competence dimensions suggested by Fiske et al. (2007; Skowronski & Carlston, 1987). Once lost, it is easier to reclaim competence than it is to reclaim warmth (Skowronski & Carlston, 1992). Indeed, if a target is deemed to be unfriendly or immoral, any attempts at being friendly or moral may be attributed to social pressure, rather than intrinsic warmth (Wojciszke, 2005). Under the cue diagnosticity model, this is because immoral or antisocial behaviour is deemed to be more diagnostic of morality than social or moral behaviour (Skowronski & Carlston, 1989, 1992). Rather simplistically,

[4]The images are typical 'stock' images, but really draw attention to the fact that nearly all stock images presenting competent people have slender or lean models.

the model states that people assume that moral individuals will only act morally; whereas immoral individuals may act immorally or morally. An observation of an immoral behaviour, therefore, will exclude a target from the set of moral people, which displays of moral behaviour cannot rectify (Reeder & Brewer, 1979; Reeder, Pryor, & Wojciszke, 1992). The same can be said for competence: a competent person may display competent or incompetent behaviours, whereas an incompetent person will only display incompetent behaviours (Skowronski & Carlston, 1992).

Negative valence asymmetry

Much work has been conducted on the so called 'valence asymmetry hypothesis'. This states that the effect of different valences on evaluations is asymmetrical because negative experiences worsen evaluations more than positive experiences improve them. A meta-analytic moderation analysis on valence asymmetry studies suggests that negative information outweighs positive information when a group is stigmatised, but that this is reversed when a group is admired (Paolini & McIntyre, 2019). According to the meta-analysis, positive information about a stigmatised out-group still improves evaluations ($r = 0.25$), however not as much as negative information decreases evaluations. In the current context, this means that if the exercise behaviour performed by the exemplar is deemed to be 'positive' information, it may increase positive evaluations of competence.

This theory is derived from evolutionary psychology[5]: dishonesty is more likely to benefit the self and not the group/others - and so once a person is deemed to be dishonest, no amount of honesty or moral behaviour[6] will affect that evaluation (Fiske et al., 2007). After being evaluated as 'cold', any amount of 'warm' demonstrations will be interpreted as attempts to benefit the self (Fiske et al., 2007).

[5] Which will only be given a cursory mention here.
[6] Or at least an unfeasibly large amount.

This is not the case with competence information. For example, upon being rated as unintelligent, frequent displays of intelligence will eventually outweigh this judgement. A potential factor in this model is that morality can be seen as directly benefitting or harming others, and therefore carries information relevant to survival; whereas competence benefits only the individual (unless partnered with morality). So, from an evolutionary perspective, to avoid harm, it is best to keep away from those who are deemed to be immoral (Peeters, 1983). In Chapters 6 and 7, larger bodied people were evaluated as friendly but incompetent, whereas athletic people were evaluated as competent but unfriendly. Following the above models of impression formation, competence evaluations of larger-bodied individuals ought to be more readily improved through the provision of contradictory information. In the current research, competence was a hybrid construct made up of general terms such as dedication, and health-specific words such as 'active'.

Exclusivity and Diagnosticity

To determine the extent to which a behaviour is diagnostic of a group, Skowronski & Carlston (1987) developed the *cue validity index*. Participants rate the likelihood that two group members with opposite traits (e.g. moral/immoral) would perform a behaviour. These probabilities are divided by the sum of the probabilities of the existing trait and the opposite trait. This index provides an estimate of the exclusivity of a behaviour to a group sharing a trait. The behaviour is completely exclusive when the index is zero (Skowronski & Carlston, 1992).

The exclusivity of a behaviour to a particular group is related to its extremity (Skowronski & Carlston, 1987). This is discussed in terms of honesty and competence. For example, extreme dishonesty (e.g. "embezzling funds that were donated for a child's liver transplant") is very diagnostic of dishonest people; whereas a minor violation (e.g. telling 'white' lies; downloading an MP3 from a peer-to-peer service) may have been performed by the honest and dishonest alike (Skowronski & Carlston, 1987, 1992). There are perhaps few single behaviours that would be extreme enough to entirely exclude a group based on body-size alone, since body-shape is stereotypically attributed to an accumulation of behaviours related to lifestyle and general self-control, rather than single acts (Puhl & Brownell, 2006). It is possible that any observation of a larger bodied person exercising would be paired with a belief that it is either irregular, or the amount of perceived unhealthy behaviours that led the person to their physique far outweigh the immediately obvious healthy behaviours. However, the Negative Valence Asymmetry allows predictions to be made about the effect of positive information on impression formation. Positive information should affect person formation, but with a smaller differential than negative information.

Relevance to Weight Stigma

Although Fiske and colleagues conceptualise the 'uncontrolled' dimension of stereotypes in terms of warmth, others have suggested a more specific 'morality' dimension (Wojciszke, 2005). Morality is related to honesty, self-transcendence, and generosity. Weight stigmatisation has been mapped onto the dimension of morality through the so-called deadly sins of sloth and gluttony. These have become embedded in discourse surrounding the 'epidemic' of obesity, as wilful transgressions (Townend, 2009). These discourses are prevalent throughout facets of the media, as reported in a content analysis of portrayals of obesity in the UK (Flint, Hudson, & Lavallee, 2016).

The competence dimension in the structure of weight stigma is related to laziness and has less of a moral element since it is often deemed to be a controllable behaviour. Using competence and morality as an experimental manipulation, Täuber, Gausel, & Flint (2018) found that when larger bodied participants were presented with vouchers focusing on the so-called controllable element of the stigma (laziness), they showed stronger intentions to perform health-related behaviours such as exercise and healthier eating than when the moral elements were the focus. The authors suggest that these findings have important implications for health promotion, and that by portraying larger-bodied individuals as immoral and greedy, there is a risk that intrinsic motivation will not be fostered.

Although focussing on incompetence as an overweight-related trait may provide people who would benefit from healthier lifestyles with intrinsic motivation, this still reinforces the myth that fitness and larger-bodies are inversely related, when there is evidence to suggest that they are in fact orthogonal (Blair et al., 1995; Lee et al., 1999). The current study aims to assess the effect of decoupling this illusory negative correlation by presenting larger bodied individuals performing fitness-related behaviours. It is hypothesised that negative competence evaluations of the larger-bodied group members will be reduced, whilst preserving the warmth-related items.

Building on Pearl et al. (2015a), in the current study, animated 3D representations will be used as exemplars of two different groups, (those with athletic physiques, and those with larger bodies), and participants will be asked to rate a list of 53 words in terms of how applicable they are for the exemplar.

The aim of the current study was to determine whether ratings of words associated with visual representations of body size could be manipulated using animations. The hypothesis was that participants would rate larger-bodied representations as fitter, but no less warm when the model is displaying health-related behaviours, compared with idle behaviours.

8.3 Method

Cue Validity

Tests of individual items did not feature in the Chapter 6 or 7 analysis plans. However, a useful check for the suitability of exercise as a stereotype-incongruent behaviour was to run an ordinal regression on the single item *exercises regularly*, with physique as a moderator. This gave a quasi-estimate of the cue validity index of exercise in the form of the change in cumulative probability of higher cut-points being selected. Results showed a dramatic (10 fold) decrease in the posterior odds of higher cut-points being selected for larger-bodied individuals (see Figure **??**).

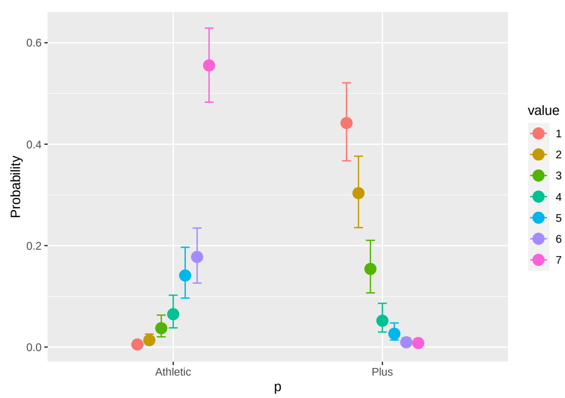


Figure 8.1: Probabilities of each of the applicability cut points for a single item (exercises regularly) being selected in by people presented with athletic or plus sized static exemplars

Although exclusivity cannot be claimed based on this observation (especially since the outcome was applicability, rather than probability), it does suggest that exercise is predominantly in the domain of athletic stereotypes. Indeed, 'very applicable' was assigned almost zero probability for the larger-bodied group. However, there was a wider distribution of probability across the lower cut-points, which broadly suggests that not exercising regularly is less diagnostic of larger-bodied people than exercising is for athletic-bodied people. A potential[7] explanation for this is that participants believed that larger

[7]Unconfirmed; this was not a hypothesised test.

bodied individuals may occasionally adopt exercise for weight management, although lack of determination prevents this from *having the desired effect*. This is particularly important because the item is conditional on exercising *regularly*, rather than occasionally. There is a suggestion in this observation that regular exercise sits towards the extreme of cue validity, but may not be sufficiently exclusive to have the nullifying effect on impression adjustment (Skowronski & Carlston, 1987, 1992).

Obviously, further work to confirm the degree to which the assumption of exclusivity of regular exercise is warranted, but this falls outside of the scope of this project.

Design

The study had a 2×2 between subjects (body-type $\times$ animation type) design. Each participant evaluated avatars of male and female genders, and both stimuli had the same animation (e.g. both running or both stationary).

Hypothesis

Assuming that the addition of animations to the exemplar will provide further information about the fitness of the exemplar, the hypotheses for this study are:

- *H1.1* Evaluations of exemplar-based competence will be higher in the animated-running condition, compared with the animated-idle condition.
- *H1.2* These evaluations will be more positive than the stationary larger-bodied group, but less positive than the active athletic group.

This is because the difference between plus and athletic body types in Chapters 7 and 6 for competence words were very large, and previous research suggests that the effect of positive information ($r = 0.25$) would not be strong enough to reverse the effect.

Analysis Plan

Data were analysed using a multi-level Bayesian ordinal generalised linear model using the R package `BRMS` (Buerkner, 2016). Priors were taken from a variety of sources, including the posterior distribution of the data from previous chapters, and previous research on impression formation. The linear equation for the regression is shown in equation (8.1)

$$y \sim \alpha + \beta_{body} : \beta_{anim} + (1|participant) \tag{8.1}$$

Hypothesis Testing

The direction and strength of the effects will be interpreted as continuous evidence, rather than using a traditional hypothesis testing method such as Bayes Factors. This will require a deeper interpretation of the results, rather than relying on cut-offs as decision rules.

Participants

Inclusion Criteria

To be eligible for this study on Prolific, participants were required to be over the age of 18, have normal to corrected-normal eyesight, and be able to understand written English. Since this study was concerned with meta-stereotypes, the inclusion criteria was kept broad, and no specific knowledge or experience was required.

Table 8.1: Age and gender information for the participants in this study.

	Female	Male	Non-binary
18 - 24	43	42	2
25 - 34	55	51	3
35 - 44	37	19	0
45 - 54	23	3	0
55 - 64	8	4	0
65 - 74	2	1	0
75 - 84	0	1	0
Prefer Not to Say	7	1	0

Sample Size

Participants were recruited using the online crowd-sourcing platform, Prolific. Participants were reimbursed the equivalent of £1.00 for completing the study[8]. Funds were acquired through a PsyPag Bursary application. In total, 302 participants were included in the analysis. A sample size of 300 respondees on the crowed sourcing platform Prolific.ac was chosen for this study, and these were recruited from Prolific.ac. This number was imposed due to limited resources available. Note again that the sample size was below the suggested minimum of 400 (Goretzko et al., 2019).

Demographics

Demographics for the sample are presented in Table 8.1. The spread of the age groups was wider for women, with larger numbers of people over the age of 35 in this group. The modal age group was the same for both genders. This sample was slightly more balanced, although in this case there was a majority of male participants. Once again, the distribution was skewed towards younger age groups.

[8] Earning an average of £7.50 per hour of participation

Apparatus

The same exemplars used in Chapter 7 were used in the current study. The avatars were animated using FBX files taken from Mixamo.com (Adobe, 2018). Two animations were downloaded, one involved the exemplar standing, idly shifting from foot to foot. The second depicted the exemplar running on the spot. Exemplars were animated in a charcoal grey room. A video clip of each exemplar (male, female; plus, athletic; running, idle) was captured using a QuickTime screen record function. The clips can be found on the OSF page for this project. The clips were sized such that they would fit on the screen with the word to be evaluated. Clips were presented using the Qualtrics survey platform (Snow & Mann, 2013). The list of 53 words identified in Chapter 6 were presented to each participant twice (two genders in the same condition), and they were asked to rate the words on the same 7 point scale in terms of applicability as in Chapter 6 and 7.

Procedure

After providing their informed consent to participate in the study, participants were assigned to one of four conditions (Plus:Athletic, Running:Idle). Each participant responded to both male and female genders, and the order of presentation was randomised. Following ratings of the words, participants rated the 53 words once more in response to the verbal descriptor "People with larger or plus-sized bodies" to test for individual-to-group transfer effects.

Participants were then presented two clips with the exemplar in a blocked random order. Underneath the clip was presented each of the 53 words found in Chapter 6 and retested in Chapter 7. In total, participants rated 106 words. The clip was set to play automatically and loop in case participants took longer to respond to the clip. Underneath the clip, each of the 53 words were sequentially presented in a random order for the participant to rate on a 7 point Likert scale

(1 = not applicable, 7 = very applicable). See Figure 8.2 for an example.

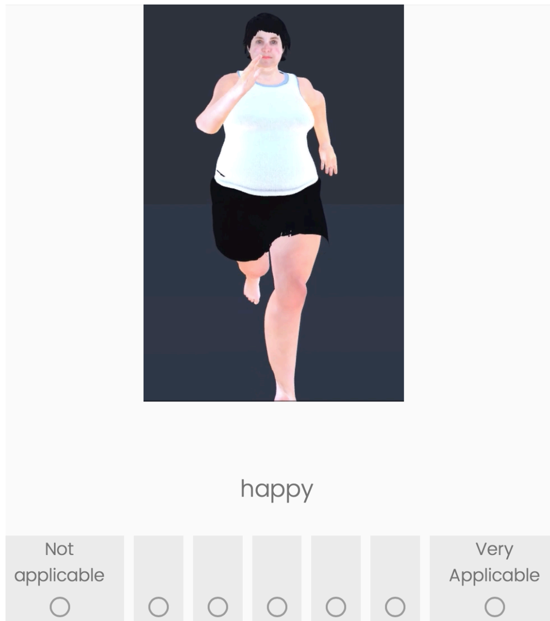


Figure 8.2: Example of active exemplar stimulus

8.4 Results

The confirmatory results section follows a pre-registered design which can be found at osf.io (https://osf.io/5arnc).

Sampling Adequacy

First exposure

Sampling adequacy for the first exposure dataset was determined using Kaiser's Little Jiffy Mk4 (AKA Kaiser, Meyer, Olkin index). The scores for the full set KMO of 0.95, for the athletic dataset KMO = 0.89, and for the plus sized dataset KMO = 0.87.

Second exposure

Sampling adequacy for the second exposure yielded a KMO of 0.96, for the athletic dataset KMO = 0.88, and for the plus sized dataset KMO = 0.88.

Both exposures The scores values for both exposures participants yielded a KMO of 0.97, for the athletic dataset KMO = 0.94, and for the plus sized dataset KMO = 0.94.

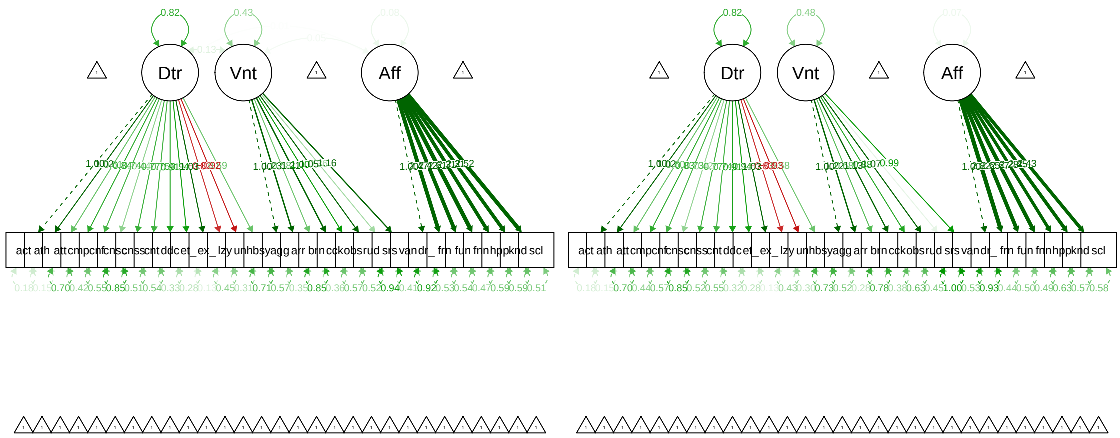
CFA

Confirmatory Factor Analysis was run on the set of 29 words from Chapters 6 and 7. The model fit was not adequate for the single (first) observations (RMSEA = 0.088). It is once again likely that there was an insufficient sample size for the full model. The full model was decomposed into three separate factors, meaning that there were fewer overall parameters to estimate. The fit indices were good for each factor (note that this was not the case for the smaller sample size, in which factor three was above the threshold of 0.08). The factor diagrams are presented in Figure 8.3a and 8.3b[9]

Regression

As in Chapters 6 and 7, ordinal probit regression was run on the visual data to determine whether the factors from the verbal phase also varied depending on the physique the exemplar. In this case, gender was excluded from the analysis in the confirmatory phase.

[9]It is important to note however that the decomposed factor model was not pre-registered and should be interpreted as exploratory. Further, the pre-registered sample size was exceeded, and the analysis was run twice, and the results of the first analysis informed the continued collection of data. Although optional stopping has not been discussed in terms of CFA (yet), it is possible that the error rates have been inflated due to this practice.



(a) Full CFA model RMSEA = 0.088

(b) Individual CFA models for each factor. They are presented on the same pane for brevity, but were not entered into the analysis together

Figure 8.3: CFA models

Prior Selection and Sensitivity Check

As before, the sensitivity of the likelihood to the prior distributions was checked by running the models multiple times with different priors. The first priors were taken from the updated posterior distribution from Chapters 6 and 7. Using the `fitdistrplus` package in R (Delignette-Muller & Dutang, 2015), the mean and standard deviations from the physique slope parameter (β_p) were estimated. Next, a prior for the effect of 'positive' information on evaluations of larger bodied people (i.e. the interaction between physique and action) was provided by adding the meta-analytically estimated effect of positive information on stigmatised groups to the effect of physique (Paolini & McIntyre, 2019). The estimated effect size reported by Paolini et al, as correlation coefficient, was $r = 0.25$, which, when converted to standardised mean effect size yielded $d = 0.67$. This was added to the posterior estimate of physique on competence evaluations from Chapter 7, yielding a prior estimate of N(0.5,5). Since there was no theoretical reason to believe that Factors 2 and 3 would be affected by the action of the exemplar, wide distributions were included as priors for these models (see Table 8.4). The second set of priors were the wide, weakly informative priors used in Chapters 6 and 7 were placed on the beta coefficients, constraining 95% of the distribution to 0 ± 10.

There was minimal variation in the posterior estimates under different prior conditions for Factors 1 and 2. For Factor 3, the estimated effect of physique was stronger under the weaker prior distribution, but the effect of action was stronger under the informative prior. There was a relatively strong reduction in competence ratings when the avatar had a larger body.

Table 8.2: Regression table of the effect of animation and physique on word rating for the three factors with weakly informative priors

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.34	-2.62 – -2.06	-1.61	-2.08 – -1.16	-1.61	-1.88 – -1.36
3	Intercept.2.	-1.61	-1.89 – -1.34	-0.93	-1.40 – -0.49	-1.09	-1.35 – -0.84
4	Intercept.3.	-0.91	-1.19 – -0.64	-0.33	-0.79 – 0.12	-0.52	-0.77 – -0.27
5	Intercept.4.	-0.13	-0.40 – 0.15	0.75	0.29 – 1.21	0.86	0.61 – 1.12
6	Intercept.5.	0.65	0.37 – 0.92	1.55	1.08 – 2.01	1.86	1.60 – 2.12
7	Intercept.6.	1.41	1.13 – 1.69	2.36	1.89 – 2.84	2.74	2.45 – 3.03
8	p	-1.56	-1.77 – -1.34	-0.35	-0.62 – -0.07	0.44	0.17 – 0.72
9	act	0.83	0.62 – 1.05	-0.18	-0.45 – 0.10	-0.18	-0.45 – 0.10
10	p.act	0.24	-0.07 – 0.54	-0.04	-0.43 – 0.35	-0.11	-0.51 – 0.28
11	Random Effects						
12	var	0.47	0.80	0.60			
13	precis	2.74	1.62	1.24			
14	ICC	0.15	0.33	0.33			
15	N	14 idx	8 idx	7 idx			
16		302 id	302 id	302 id			
17	Marginal R2 / Conditional R2	0.342 / 0.531	0.029 / 0.415	0.031 / 0.369			

Note.

Table 8.3: LOOCV IC estimates for the three models

	LOOIC	SE
F1 ICP-PHY	17.72	9.19
F1 ICP-FUL	37.03	12.60
F1 P-FUL	19.31	8.65
F2 ICP-PHY	1.50	3.39
F2 ICP-FUL	1.99	4.95
F2 P-FUL	0.49	3.31
F3 ICP-PHY	2.87	2.55
F3 ICP-FUL	3.12	3.55
F3 P-FUL	0.24	2.38

Model Selection

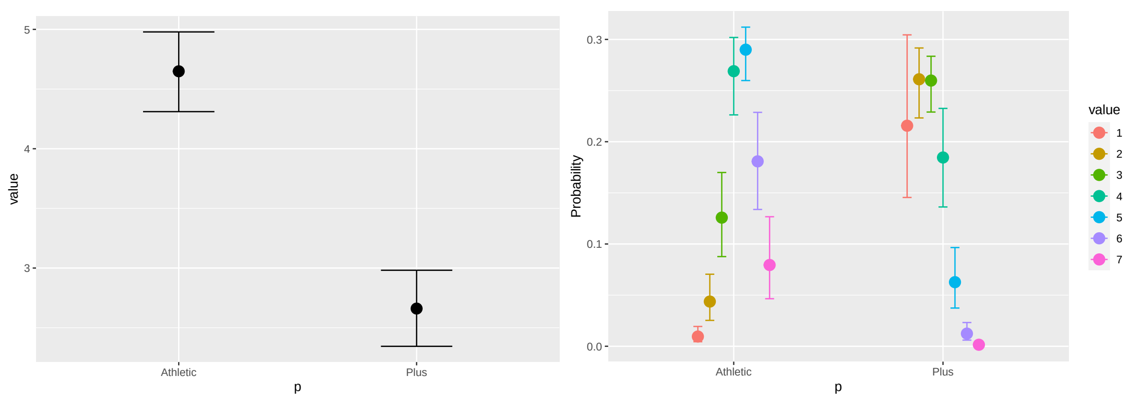
For each factor, Leave One Out Cross Validation (Vehtari et al., 2018) was applied to the three models (intercept, physique, all factors). The differences can be found in Table 8.3. In all cases, the full model offered the better fit, however this was overwhelmingly evident in Factor 1 in which the difference between the intercept and full model LOOIC was nearly three standard errors in magnitude. That is, including the type of animation of the exemplar to the Factor 1 model improved the fit of the data, relative to the complexity that it added. It is again important to note that the differences in LOO IC were not particularly large for Factor 2 and 3. The full models for each factor will be used in the following results section.

Model Descriptions

The direction and strength of the effects of physique found in Chapter 6 and 7 were broadly replicated in the animated avatar study (although a direct comparison will be made in the exploratory section). In the following sections, Factor 1 will be addressed in isolation as the pre-registered hypothesis test. Next Factors 2 and 3 will be discussed.

Table 8.4 provides the parameter estimates for each factor in the full model (p*g).

Factor 1 (Pre-registered Confirmatory analysis) Once again, under the full model, applicability ratings of the Factor 1 words were the most affected by the physique of the exemplar. The mean effect of physique on applicability ratings of these words was reduced by β = -1.60 standard deviation points on the hypothetical latent distribution. The 95% credible interval for this parameter was 95%CI = [-1.72 – -1.47], meaning that, given the prior information and the current data, there is a 95% probability that this physique has a large negative effect on evaluations of competence-related words (see Figure 8.4a and 8.4b.

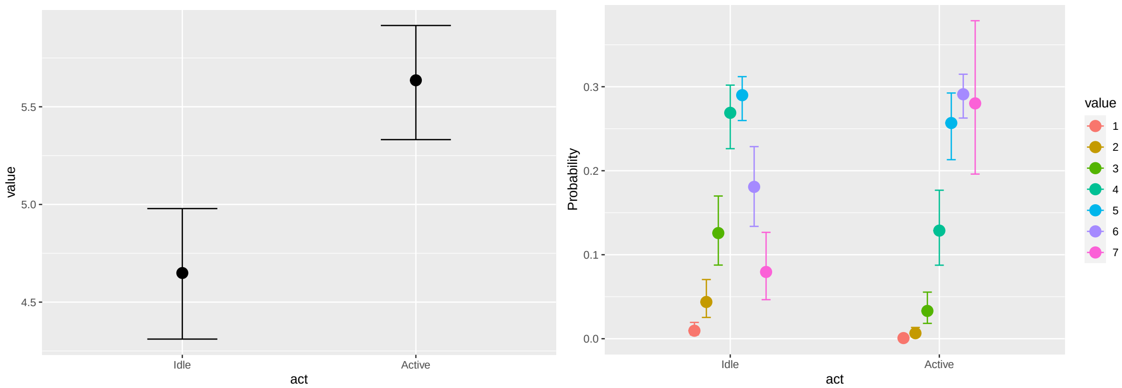


(a) Marginal effects plots for the effect of exemplar physique on Factor 1 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar physique on Factor 1 word rating task.

Figure 8.4: Marginal plots for the effect of exemplar physique on Factor 1 applicability ratings

The action of the avatar had a large positive effect on the competence ratings of the exemplars. The mean effect of action on applicability ratings of these words was reduced by $\beta = 0.81$ standard deviation points on the hypothetical latent distribution. The 95% credible interval for this parameter was 95%CI = [0.60 – 1.01], meaning that, given the prior information and the current data, there is a 95% probability that this the action of the exemplar has a large negative effect on evaluations of competence-related words (see Figure 8.5a and 8.5b).



(a) Marginal effects plots for the effect of exemplar physique on Factor 1 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar animation on Factor 1 word rating task.

Figure 8.5: Marginal effect plots for the effect of exemplar animation on Factor 1 word ratings

There was also an interaction between physique and action on competence evaluations of the exemplars. This was in the hypothesised direction. When the exemplar had a plus-sized body, and was running on the spot, there was a larger increase in the evaluations of competence, compared with the athletic exemplar. The interaction estimate for Factor 1 was 0.28 standard deviation points on the hypothetical latent distribution. The 95% credible interval for this parameter was 95% CI = [0.04 – 0.54] (see Figure 8.6). In sum, the results favour the hypothesis that additional, counter-stereotypical information would improve competence evaluations of larger bodied exemplars more than athletic-looking exemplars, although evaluations were higher across all conditions.

Factors 2 and 3 Although there was no hypothesis attached to Factors 2 and 3, exploratory analyses were run on these factors anyway. As in Chapters 6 and 7, there was a similar reduction in arrogance ratings (Factor 2) and an increase in friendliness ratings (Factor 3) for larger bodied exemplars, compared with athletic exemplars.

For Factor 2, the physique condition was associated with a reduction of β = -0.30 standard deviation units on the latent scale. The 95 % credible interval

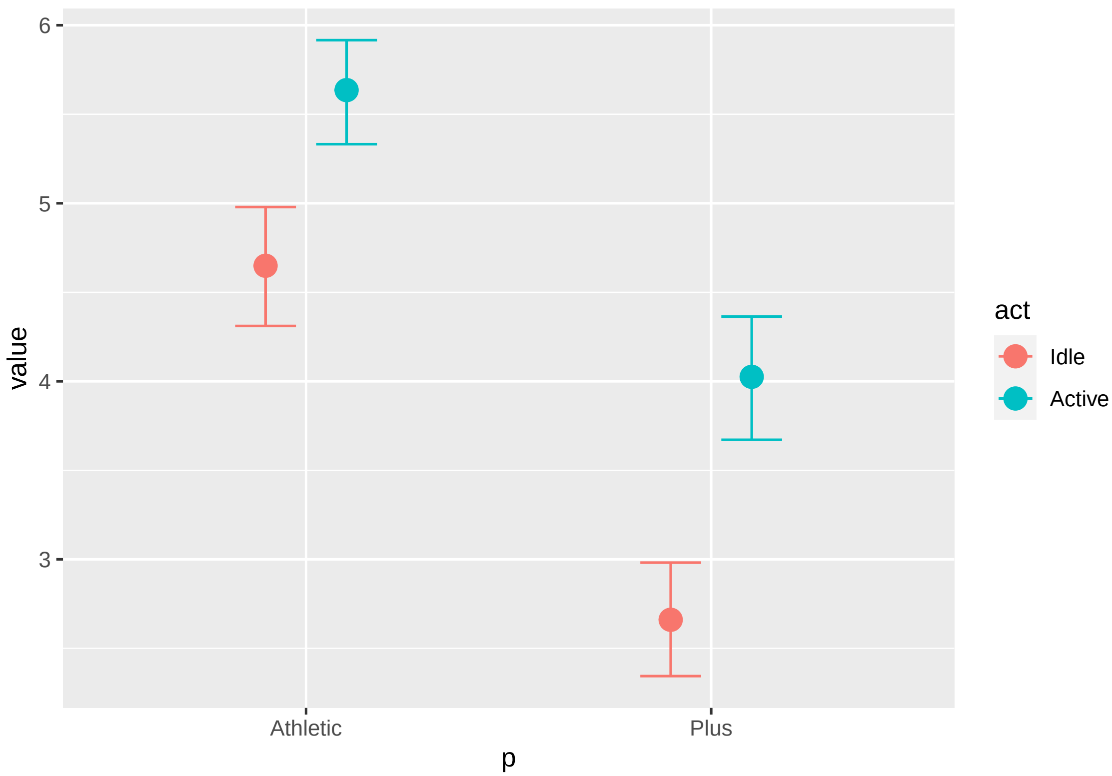
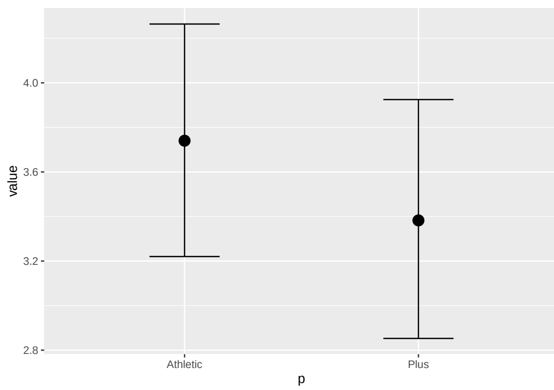
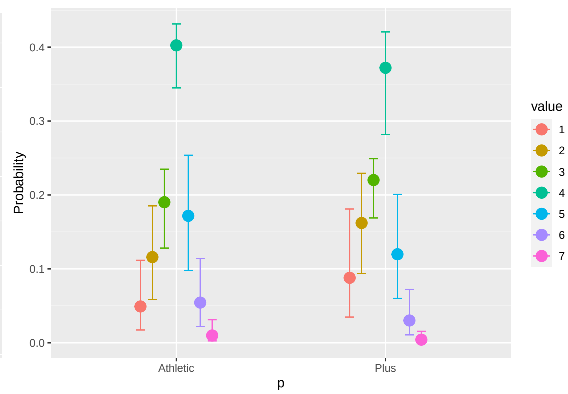


Figure 8.6: Marginal effects plots for the interaction between exemplar physique and animation on Factor 1 word ratings.

was 95%CI = [-0.44 – -0.16], meaning that we can be very confident that the effect of plus-sized bodies on these evaluations is negative and moderately sized (see Figure 8.7a and 8.7b).



(a) Marginal effects plots for the effect of exemplar animation on Factor 1 word ratings.



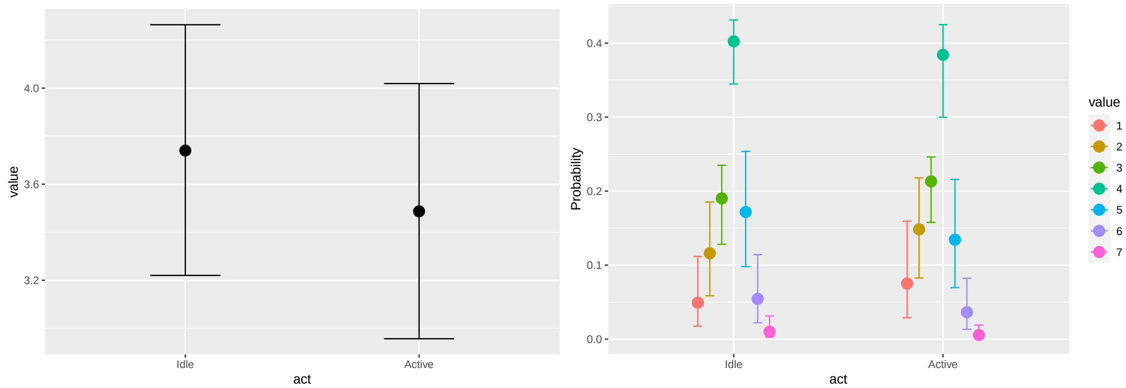
(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar animation on Factor 1 word rating task.

Figure 8.7: Marginal effect plots for the effect of exemplar physique on Factor 2 word ratings

The action condition had a similarly sized, but less confident effect on arrogance-related word applicability ratings, with reduction of β = -0.21 standard deviation units on the latent scale. The 95 % credible interval was 95%CI = [-0.35 – -0.07]. This suggests that running avatars were rated as less arrogant across all conditions (see Figure 8.8a and 8.8b).

The interaction between action and physique was of the same magnitude with a similar degree of uncertainty $\beta_{p.act}$ = -0.21, 95%CI = [-0.37 – -0.05] (see Figure 8.9).

For Factor 3, there was an increase in applicability evaluations of words related to Affability when the exemplar had a larger body. This also replicates the effect from previous Chapters. There was an increase of β_p = 0.23 standard deviation units in the applicability evaluations of these words. The 95%CI was [0.08 – 0.38]. The lower tail of the 95% CI was very close to zero, meaning that an effect of zero is plausible, but unlikely value, since the majority of the probability density (i.e. the mean of the posterior distribution) suggests that a moderate effect



(a) Marginal effects plots for the effect of exemplar animation on Factor 2 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar animation on Factor 2 word rating task.

Figure 8.8: Marginal effect plots for the effect of exemplar action on Factor 2 word ratings

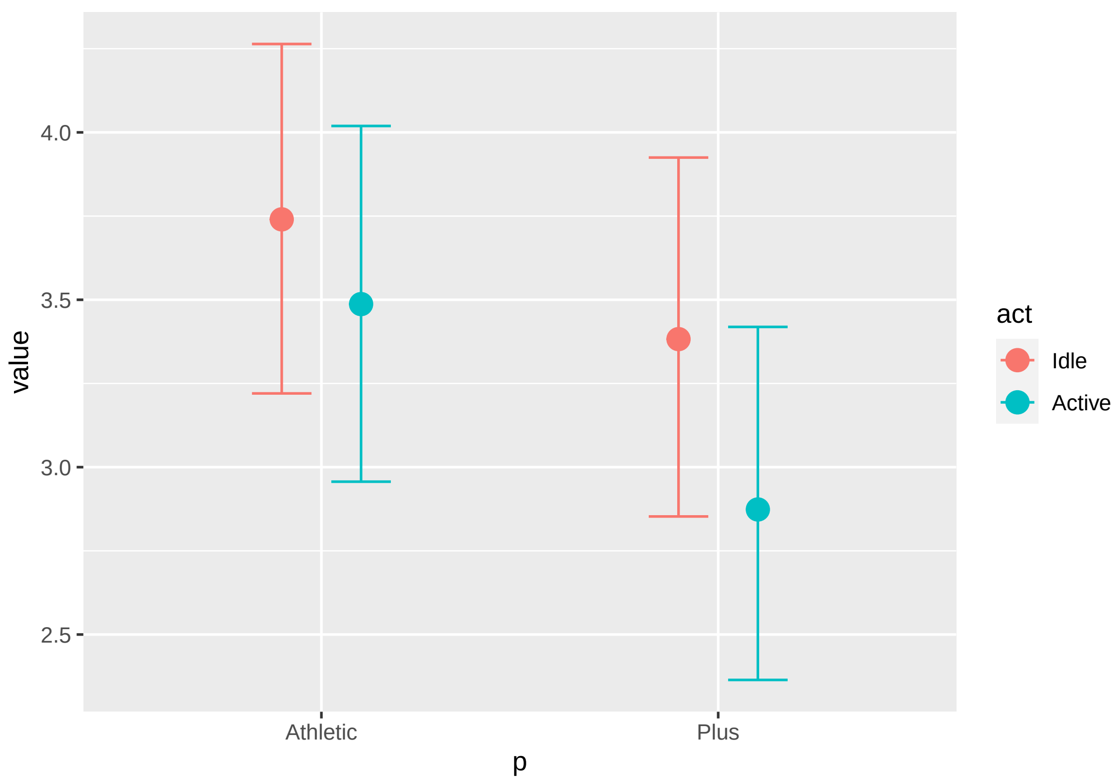
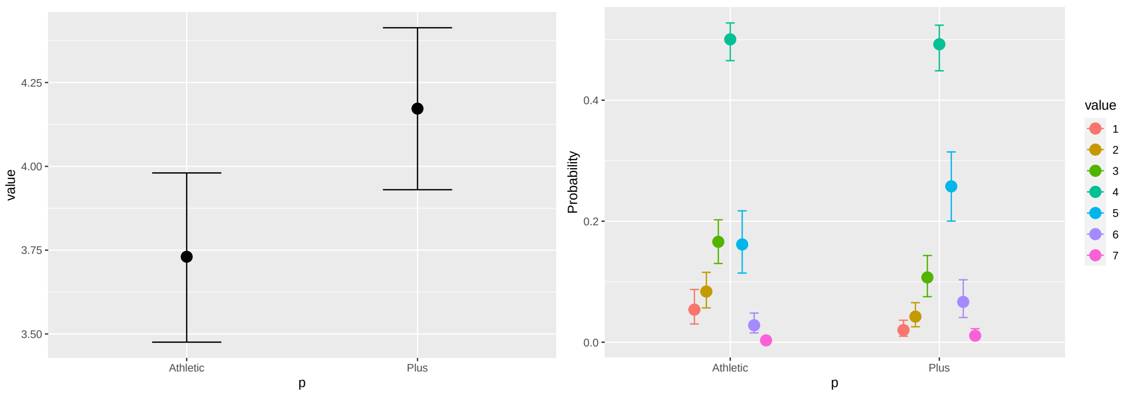


Figure 8.9: Marginal effects plots for the interaction between exemplar physique and animation action on Factor 2 word ratings.

is more likely (see Figure 8.10a and 8.10b).

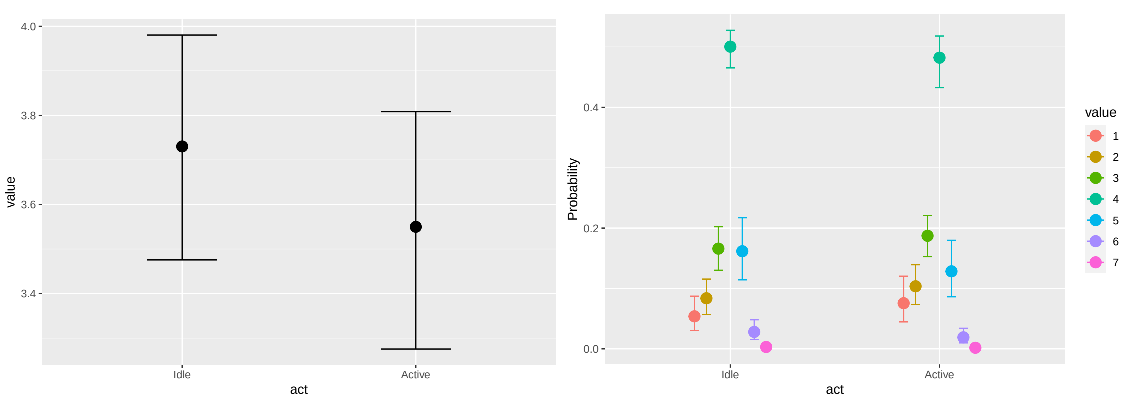


(a) Marginal effects plots for the effect of exemplar animation on Factor 3 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar animation on Factor 3 word rating task.

Figure 8.10: Marginal effect plots for the effect of exemplar physique on Factor 3 word ratings

There was a negative effect of activity on applicability ratings of friendliness words, $\beta_{p.act} = 0.10$, 95%CI = [-0.21 – 0.42], and again the lower tail of the 95% CI was very close to zero, meaning that an effect of zero is plausible. (see Figure 8.11a and 8.11b.



(a) Marginal effects plots for the effect of exemplar animation on Factor 3 word ratings.

(b) Marginal probabilities for each of the 7 Likert points for the effect of exemplar animation on Factor 3 word rating task.

Figure 8.11: Marginal effect plots for the effect of exemplar action on Factor 3 word ratings

The interaction between physique and action was minor for Factor 3, with a large degree of uncertainty, and values of and around zero were assigned higher probability, $\beta_{p.act} = 0.10$, 95%CI = [-0.21 – 0.42], meaning that we cannot be confident that there is an effect. The asymmetrical distribution suggests that most probability is assigned to positive values, so the Factor 3 words were rated higher for larger bodied running exemplars, but with the uncertainty and small effect size this is worth only a cursory mention. (see Figure 8.12).

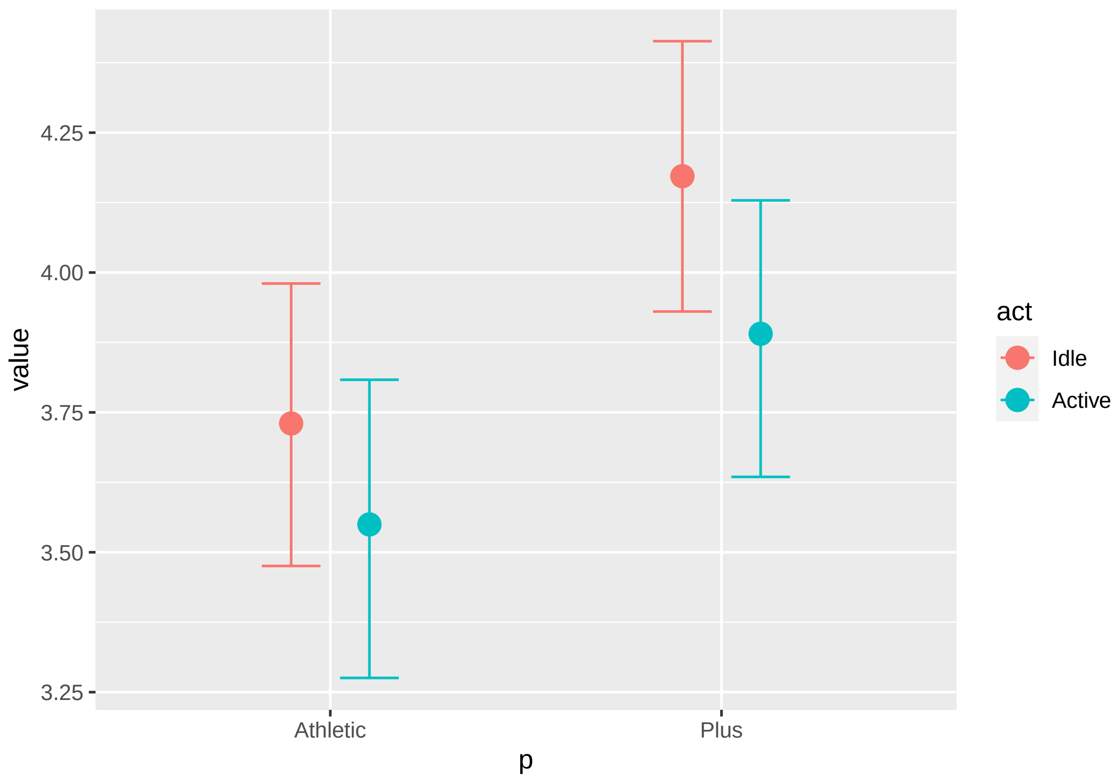


Figure 8.12: Marginal effects plots for the interaction between exemplar physique and animation action on Factor 3 word ratings.

Table 8.4: Regression table of the three Factors under the posterior distributions from the parameter estimates in Chapter 7

	Parameter	F1 Estimate	F1 CI (95%)	F2 Estimate	F2 CI (95%)	F3 Estimate	F3 CI (95%)
2	Intercept.1.	-2.37	-2.63 – -2.11	-1.65	-2.10 – -1.19	-1.71	-1.94 – -1.49
3	Intercept.2.	-1.64	-1.90 – -1.38	-0.97	-1.42 – -0.52	-1.20	-1.42 – -0.97
4	Intercept.3.	-0.94	-1.20 – -0.69	-0.37	-0.82 – 0.09	-0.62	-0.84 – -0.40
5	Intercept.4.	-0.15	-0.40 – 0.10	0.72	0.27 – 1.18	0.76	0.54 – 0.98
6	Intercept.5.	0.62	0.37 – 0.88	1.52	1.07 – 1.98	1.76	1.53 – 1.99
7	Intercept.6.	1.39	1.13 – 1.65	2.33	1.87 – 2.80	2.63	2.38 – 2.90
8	p	-1.60	-1.72 – -1.47	-0.30	-0.44 – -0.16	0.23	0.08 – 0.38
9	act	0.81	0.60 – 1.01	-0.21	-0.35 – -0.07	-0.28	-0.54 – -0.02
10	p.act	0.28	0.04 – 0.54	-0.21	-0.37 – -0.05	0.10	-0.21 – 0.42
11	Random Effects						
12	var	0.45	0.78	0.62			
13	precis	2.75	1.65	1.22			
14	ICC	0.14	0.32	0.34			
15	N	14 idx	8 idx	7 idx			
16		302 id	302 id	302 id			
17	Marginal R2 / Conditional R2	0.347 / 0.532	0.042 / 0.417	0.021 / 0.367			

Table 8.5: RMSEA, SRMS, and TLI for each model

	RMSEA	lower	upper	confidence	RMSR	TLI
Full data set	0.04	0.03	NA	0.90	0.03	0.94

EFA

As a sanity check, EFA using parallel analysis was run on the visual dataset to ensure that the factor solution identified in Chapter 6 actually replicated. This was conducted on both subsets of exposures 1^{st} and 2^{nd}.

The factors from Chapter 6 were broadly replicated, which is surprising given that the CFA model had convergence issues and a solution could not be found

Next, parallel analysis was run on the second exposure to determine whether there was within-subject variability in the factor loadings.

Results Summary

Once again, the factor structure identified in Chapter 6 was weakly supported by the CFA and replicated by the EFA methods. In this study, both the physique, and the animation of a visual exemplar was found to affect the ratings of Determination words. The athletic exemplars were rated as more determined than the plus-sized exemplars, and the exemplars that were depicted running on the spot were rated as more determined than those who were standing idly. There was also weaker additive interaction between physique and animation, with the animation of the avatar affecting the plus sized exemplars to a greater extent than the athletic exemplars. There was much overlap between the distribution estimates of Factors 2 and 3, with the mean values suggesting that idle exemplars were rated as more arrogant and friendlier than the active exemplars. With the degree of overlap between these observations, there is little to suggest that this is a robust effect.

Table 8.6: Item loadings for the three factors for the first exposure.

Item	Factor 1	Factor 2	Factor 3	NA
active	0.91	-0.05	-0.00	0.04
aggressive	0.09	0.69	-0.08	-0.04
arrogant	0.06	0.76	-0.01	-0.06
athletic	0.92	0.09	-0.05	-0.18
attractive	0.51	-0.01	0.36	-0.21
boring	-0.21	0.54	-0.09	0.25
busy	0.43	0.19	0.06	0.26
cocky	0.16	0.71	0.05	-0.12
committed	0.71	-0.09	0.13	0.35
competitive	0.71	0.25	0.07	0.08
confident	0.60	0.09	0.28	0.07
conscientious	0.25	0.07	0.18	0.36
consistent	0.61	0.11	0.10	0.17
controlled	0.61	0.18	-0.05	0.30
dedicated	0.76	-0.10	0.08	0.30
determined	0.71	-0.08	0.10	0.34
disciplined	0.81	0.10	-0.02	0.20
drinks_alcohol	-0.48	0.33	0.13	0.13
driven	0.69	0.00	0.05	0.27
eats_healthily	0.80	0.17	0.03	-0.15
energetic	0.84	-0.02	0.09	0.06
enthusiastic	0.43	-0.11	0.40	0.11
exercises_regularly	0.90	0.11	-0.05	-0.11
extroverted	0.20	0.29	0.45	-0.16
fit	0.89	0.08	-0.07	-0.19
focussed	0.65	-0.12	0.07	0.41
friendly	-0.18	-0.13	0.69	0.13
fun	-0.06	-0.11	0.69	-0.05
funny	-0.19	0.14	0.63	-0.10
goal_directed	0.72	-0.02	0.05	0.29
greedy	-0.27	0.49	0.11	0.02
happy	0.20	-0.00	0.57	-0.07
health_conscious	0.86	-0.04	0.08	0.00
healthy	0.91	0.04	-0.04	-0.15
inactive	-0.88	0.18	0.12	0.06
introvert	-0.47	0.30	0.02	0.36
kind	-0.26	0.06	0.55	0.09
lazy	0.78	-0.24	-0.16	-0.02
loud	-0.13	0.42	0.38	-0.18
moody	-0.19	0.60	-0.03	0.15
motivated	0.79	-0.12	0.13	0.21
obsessive	0.16	0.52	0.02	0.15
organised	0.35	0.16	0.32	0.16
rude	-0.07	0.72	0.04	-0.01
sedentary	-0.78	0.09	0.13	0.12
serious	0.28	0.03	0.05	0.29
social	0.11	0.11	0.64	-0.17
sporty	0.92	0.06	-0.08	-0.03
strict	0.56	0.33	0.07	0.18
strong	0.70	0.10	0.05	-0.07
unhappy	-0.49	0.38	-0.04	0.17
unhealthy	0.86	-0.06	-0.09	-0.27
vain	0.20	0.59	0.14	-0.03

Table 8.7: Item loadings for the three factors for the second exposure.

Item	Factor 1	Factor 2	Factor 3	NA
active	0.49	0.51	-0.01	-0.07
aggressive	-0.06	0.12	0.80	-0.04
arrogant	-0.02	0.02	0.83	-0.06
athletic	0.26	0.71	0.12	-0.02
attractive	-0.02	0.60	0.10	0.34
boring	0.21	-0.41	0.47	-0.07
busy	0.55	-0.11	0.07	0.14
cocky	-0.01	0.12	0.79	0.03
committed	0.85	0.02	-0.08	-0.02
competitive	0.54	0.26	0.24	0.06
confident	0.46	0.30	0.06	0.11
conscientious	0.65	-0.15	0.01	0.15
consistent	0.67	0.13	0.03	0.11
controlled	0.67	0.06	0.01	0.09
dedicated	0.80	0.11	-0.02	-0.04
determined	0.80	0.11	-0.05	0.02
disciplined	0.74	0.19	-0.03	0.03
drinks_alcohol	-0.01	-0.53	0.28	0.15
driven	0.87	-0.02	-0.02	-0.07
eats_healthily	0.25	0.66	0.15	0.13
energetic	0.55	0.38	0.07	0.01
enthusiastic	0.45	0.16	0.00	0.33
exercises_regularly	0.33	0.63	0.14	-0.04
extroverted	-0.02	0.30	0.43	0.42
fit	0.25	0.73	0.10	-0.05
focussed	0.82	0.06	-0.10	-0.03
friendly	0.05	-0.11	-0.15	0.76
fun	0.02	-0.02	-0.02	0.67
funny	-0.05	-0.11	0.11	0.66
goal_directed	0.78	0.12	-0.06	0.05
greedy	-0.03	-0.28	0.52	0.06
happy	0.12	0.24	0.00	0.60
health_conscious	0.52	0.45	0.03	-0.05
healthy	0.21	0.72	0.10	0.06
inactive	-0.21	-0.69	0.02	0.08
introvert	0.29	-0.65	0.12	0.12
kind	0.14	-0.23	-0.08	0.65
lazy	0.26	0.64	-0.18	-0.17
loud	-0.06	-0.16	0.63	0.21
moody	0.08	-0.42	0.42	0.15
motivated	0.71	0.21	-0.08	0.08
obsessive	0.49	-0.20	0.47	-0.06
organised	0.63	-0.04	0.00	0.26
rude	-0.17	-0.02	0.77	-0.03
sedentary	-0.20	-0.58	0.06	0.22
serious	0.62	-0.29	0.10	0.06
social	-0.03	0.18	0.20	0.58
sporty	0.30	0.68	0.07	0.00
strict	0.69	0.02	0.21	-0.04
strong	0.38	0.36	0.10	0.00
unhappy	0.06	-0.62	0.40	-0.02
unhealthy	0.06	0.81	-0.04	-0.03
vain	0.09	0.15	0.59	0.09

8.5 Discussion

Similar effects of physique to those found in Chapters 6 and 7 were observed in the current study. This is a promising finding, since it suggests that regardless of modality (e.g. internal representation primed by a verbal cue, visual, video), similar evaluations are made for larger bodied and athletic exemplars. The magnitude of the differences remained similar across the different modalities, with the effect of physique on competence ratings being the largest by far. If this adequately represents stereotypes associated with larger bodied individuals, and the Proteus Effect holds, the findings from previous studies are unsurprising (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014). What is surprising is the large effect of behavioural information, and the moderate interaction between physique and behavioural information in the Factor 1 ratings. This suggests that at least a portion of the negative stereotype construct associated with larger bodied individuals may be negated by providing stereotype-incongruent information. Further, the minor effect of behaviour on arrogance evaluations across conditions might be explained by a reduction in the perceived 'natural ability', i.e. the exemplar may affect authentic pride rather than hubristic pride when exercising (Tracy & Prehn, 2012). The minor reduction in the perceived 'friendliness' of exercising exemplars is also interesting. They may be perceived as boring, more focussed on their health, and less likely to socialise. This observation was not supported by an interaction, suggesting that it was a main effect of activity. In the absence of a 'stigmatised' condition (i.e., providing stereotype congruent information such as lounging or eating) for the larger bodied exemplars it is difficult to draw firm conclusions, however the current findings suggest that presenting a larger bodied model being active reduces negative evaluations of competence. This both supports and contrasts the findings of Pearl et al. (2015a), who concluded that neutral settings were the least stigmatising. In this study, the neutral pose elicited a similar difference between athletic and plus-sized ratings as in Chapters 6 and 7, whereas the exercising larger bodied exemplars were rated as more competent. It must be noted that the measure used by Pearl et al. (2015a) (a 14 item 'Fat Phobia' scale), are somewhat different. The Fat Phobia scale measures negative stereotype-endorsement:

> Directions: Listed below are 14 pairs of adjectives sometimes used to describe obese or fat people. For each adjective pair, please place an X on the line closest to the adjective that you feel best describes your feelings and beliefs. (Bacon et al., 2001, p. p257)

The current study asked participants to rate 'out there' stereotypes, rather than 'in here stereotypes' (Kowert et al., 2012). The distinction is that the former asks about the beliefs about the presence of a stereotypical belief in the general population between, and the latter asks about personally held beliefs (i.e. prejudice when negative). These are very different questions. For 'in here' stereotypes, the subjective belief has already been formed, and may be susceptible to social desirability bias. It was noted earlier that there are similarities between the fat phobia construct and the set of words developed over the past three Chapters, and there may be some overlap in the content space. However, the current set of words also contains positive elements, meaning that the ambivalent structure of evaluations could be addressed. An interesting construct validation study would be to correlate the negative attitudes towards larger bodied people with the applicability ratings of this set of words. If the two stereotype domains (in and out) are independent, then there ought not be a difference between applicability ratings of the negative words. However, if the wording of the fat phobia scale is changed to the out-domain, there would be positive correlation with the competence ratings, and a negative correlation with the arrogance and friendliness ratings.

As suggested by McIntyre, Paolini, & Hewstone (2016), repeated positive exposure to individual members of a group may result in positive group evaluations (i.e. a reduction in 'in here' stereotypes about groups). However, a positive exemplar must first be defined. This study has demonstrated that 'out there' stigmatising stereotypes may be reduced by including counter-stereotypical information, a further step, in addition to the work by the Rudd Centre and Pearl et al. (2015b), towards improving group evaluations.

Link to the Proteus Effect

The animated exemplars evoked similar stereotype structures to the prototypical descriptors and the static images representing larger bodied and athletic groups. The Proteus Effect requires that avatars carry this information in order for behavioural assimilation to occur. This study has established that perceived *out there* stereotypes may be influenced through the introduction of additional behavioural information in a way predicted by Associated Systems Theory, and Negative Valence Asymmetry theories. Given how perceived general attitudes towards larger bodied people may be altered through this behavioural information, the Proteus Effect theory would predict that the assimilation of lazy behaviour would be reduced in players who are presented with an active larger-bodied avatar. This prediction is testable and will be addressed in Chapter 9.

Strengths and Limitations

A strength of this study (and indeed the previous study) was that the sample was not drawn from an exclusively student population. A range of age groups and genders participated. However, race and ethnicity data were not collected and so it is unclear how diverse the sample is along these dimensions.

A second strength is the within subject, randomised design. A lesson learned from Chapter 7 was that presenting the same gender avatar across both physique conditions may have biased the second exposure results. On this occasion, participants responded to the same physique and action condition, but different genders, meaning that potential differences in the ratings of same and different gender could be explored.

A limitation of this study is the absence of a negative stereotype congruent animation. This would have allowed for a more in-depth exploration into the linearity of the congruence effect, as well as an interesting test of the Negative Valence Asymmetry hypothesis. The inclusion would, however, have added further complexity to an already complicated study, and would be best added to a future study. The unbalanced gender grouping may have been an additional issue, since the age variance appeared to be larger in the female group than the male group. Once again, it is still unclear whether meta-stereotypes vary between groups. Cuddy et al. (2009) suggest that the warmth competence dimensions are universal across cultures.

Applied Context

Static visual portrayals of larger bodied people generally have an impoverished information content. This was shown in Chapter 7, and also to some extent in the work of Pearl et al. (2015a) who demonstrated that simply presenting larger bodied people in exercise contexts was not sufficient to reduce fat-phobia related evaluations; whereas negative evaluations were reduced when the context implied higher (e.g. managerial) status. This chapter has demonstrated that additional behavioural information may reduce negative evaluations, which may be used by organisations such as the Rudd Centre to promote positive (or at least less negative) depictions and evaluations of larger bodied people. For example, promotional materials utilising behavioural information may be used in health promotion to both reduce the negative self-evaluation that larger bodied people may feel about their body size and the appropriateness of exercise (cf. Chapter 2, page 47). Moreover, by presenting larger bodied individuals as unambiguously competent in an exercise context, the degrees of freedom available to observers to make negative evaluations are reduced; and moreover, widespread positive representation may have the cumulative effect of reducing *out there* stereotypes, as suggested by Puhl et al. (2013).

Future Directions

The next step in this research programme is to determine whether the observed effects of physique on attitudes translate into behaviour during an exercise videogame. However, this study itself could be developed further by including a 'stigma-congruent' condition to determine where the baseline applicability ratings for the three factors sit. It would also be interesting to see whether the other factors can be manipulated to a similar extent as the competence dimension using factor specific positive and negative behavioural information. As with studies on both sides of this one (Chapters 7 and 9), changing avatar characteristics further would be interesting. For instance, no intermediate exemplar (i.e. neither muscular nor plus sized) was included as a stimuli. It would be interesting to check whether the effect of activity on applicability ratings was general to other body types, or merely the larger bodied representations. Further attributes such as race, age, and height may also be manipulated.

Findings and Conclusion

In summary, Chapter 6 and Chapter 7 have shown that the first two necessary conditions discussed on page 52 are met by the groups 'athletic' and 'plus-sized' people, and the exemplars that were created in Chapter 6. Based on this, we may conclude that if the Proteus Effect holds then, there ought to be a behavioural response to the $Activity|BodySize$ activation elicited by avatars (page 30). This chapter has demonstrated that this stereotype may be altered at the impression formation stage by providing additional information, in line with the predictions of Associated Systems Theory (Chapter 2, page 41) that stereotypes are resorted to when there is impoverished information (in this case, visual) available. We can therefore extend the notation to $Activity|Size \times Animation$ in the current chapter.

There were large differences in competence evaluations between larger bodied and athletic-looking exemplars across all three studies, and these appear to be adequately evoked by both visual and discursive stimuli. This suggests that if these exemplars do not affect participants performance in an exergame, this would bring into question previous research, since such a strong stereotype ought to lead to behavioural assimilation. Confidence in the transfer of Arrogance and Affability factors to behaviour is lower, since the effects are a fraction of the magnitude of the competence dimension. However, behaviour related to arrogance will still be addressed in Chapter 9 using a base-rate neglect paradigm as a proxy. In the concluding study, it will be possible to compare the strength of evaluations to the strength of the behavioural effect. This will be a first in Proteus Effect research.

Chapter 9

Exemplars, Exercise, and Exergames: A replication-extension of Li, Lwin & Jung (2014)

9.1 Rationale

In the final empirical chapter, the findings from Chapters 4 to 8 will be synthesised and applied to a replication-extension of the study by Li et al. (2014). Armed with a critical and holistic picture of the existing theory (Chapter 4); knowledge about avatar design preferences and the experience of exercising *as* a muscular heroic avatar (Chapter 5); data about stereotypes associated with both athletic and plus-sized prototypes (Chapter 6), exemplars (Chapter 7); and a means of interrupting these stereotypes using additional behavioural information (Chapter 8), the extent to which the observed appearance and behaviour of an avatar can affect exercise intensity will be addressed. That is to say are the visual properties of avatars pervasively persuasive? A novel appendix to the source study is the inclusion of a base-rate neglect paradigm to identify whether the use of an athletic avatar evokes the 'arrogance' side of the athletic stereotype. It is through this chapter that the connection between avatars and health-related behaviours will be made.

9.2 Introduction

In this Chapter, the accumulated knowledge about athletic and plus sized stereotypes from the previous three Chapters will be applied to an exergaming context in the form of a replication-extension of Li et al. (2014). Returning to the Proteus Effect, the appearance of an avatar has been shown to modify the behaviour of a user (Yee et al., 2009). This is because inferences taken from salient information about the avatar are temporarily assimilated into the user's self-perception.

In Chapter 4, evidence that the embodiment of an overweight avatar may affect in-game performance whilst playing an exergame was reviewed. The three studies demonstrating this effect assigned people either an 'overweight/obese', or a 'normal' avatar (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014). All three studies used the Nintendo Wii console, which has a motion sensing controller ('WiiMote') allowing users to control the game using larger movements. The participants in the Li et al study were "overweight or obese"" children, whereas the participants in the two Peña et al studies were college age adults who were not selected for body type (although BMI was used as a covariate). Measures of performance included the number of wrist and waist movements using accelerometers (Peña et al., 2016; Peña & Kim, 2014), or the time taken to navigate an avatar around an athletics track by running on the spot (Li et al., 2014). Applied to these previous studies, the Proteus Effect theory predicts that by adopting a larger body, participants perceived themselves to be lazier, and move less as a result.

Chapter 4 also outlined three necessary conditions for the Proteus Effect to work; existence of a visual stereotype; unambiguous representation of this stereotype by the avatar; accessibility of the stereotype to the participant. The effect of ambivalent stereotypes on behaviour has not been investigated in previous Proteus Effect studies.

There is wide and varied support that laziness forms part of the stereotype structure for larger-bodied people, and so the prediction that, if the Proteus Effect holds, less movement would be observed prediction is justified (Flint & Reale, 2018; Greenleaf et al., 2019; Puhl & Brownell, 2006; R. Puhl & Suh, 2015). The existing structure of larger-bodied stereotypes was also supported by evidence from Chapters 6, 7, and 8, in which stereotype-related words were generated and rated in response to group descriptors, images, and videos. In addition to corroborating the suggested stereotypical representations (i.e. that larger people are perceived as lazy and incompetent), an ambivalent structure for both plus sized and athletic people was found. That is, where people with larger bodies were evaluated as lazier and less dedicated, they were also rated as more friendly and less arrogant than people with athletic physiques. The message here is that, despite the presence of weight stigma, there are also positive elements associated with larger bodies. So far I have shown that negative stereotypes are unambiguously associated with larger bodies, and in two studies these were evoked by the visual exemplars. Evidence that these negative evaluations are diffuse and widely accessible is supported, not just by the replication of the ratings in three independent samples, but also the almost ubiquitous global prevalence of weight stigma, suggesting that it is likely to be available and accessible to a wide range of people (Brewis, 2014). In addition to these three conditions, I have also demonstrated that stigmatising evaluations may be partially negated by presenting the exemplar with counter-stereotypical information (Chapter 8). Indeed, participants who were presented with a larger bodied exemplar animated to be running on the spot rated the negative stereotype words as less applicable than when the exemplar was standing stationary. Taken together, a solid inference about behavioural responses to the avatars may be made. The remainder of this introduction will unpack the ambivalence of stereotypes.

Ambivalent Stereotypes Revisited

In the previous empirical studies in this PhD programme, there was evidence of ambivalent stereotype structures for both athletic and plus-sized people. Athletic people were considered healthy and competent but arrogant, and plus-sized people were considered unhealthy and incompetent but nice. One of the aims of the current study is to explore this ambivalence in reference to the Proteus Effect. In addition to replicating the study by Li et al. (2014), in which larger bodied avatars were associated with slower virtual running track completion, the potential negative 'arrogance' dimension of the athletic stereotype, and its effect on self-perception, will be explored in the current study.

In this case, base-rate neglect will be used as a proxy for *inflated self-regard*. Base rate neglect is a heuristic that is adopted by System 1 (Chapter 2, page 56), and describes the tendency to ignore factual, statistical information in favour of less relevant information. For example, when asked to guess a fictional person's job, people are more likely to choose librarian than engineer when told that the person is also a feminist, focussing on the irrelevant stereotypical information, and ignoring the base rate that there are more engineers than librarians (Kahneman & Tversky, 1973; Pennycook, Trippas, Handley, & Thompson, 2014; Thompson & Pennycook, 2016). In the current study, base rate information will be provided about a task that participants will complete. They will be asked to report how confident they are that they will achieve a goal, in lieu of this information. If using an athletic avatar increases 'arrogance', then participants in this condition ought to overestimate the likelihood that they will achieve the goal, when the task is expected to be harder. Moreover, is expected that the use of an athletic avatar will lead participants to neglect information implying a low base-rate for success, since they will overestimate their own ability. Conversely, it is expected that the use of a larger bodied avatar will lead participants to neglect information implying a high base-rate for success, since they will underestimate their own ability.

Counter-stereotypes

There is a body of evidence suggesting that stereotypes may be unlearned with the provision of anti- or counter-stereotypical information (Dunaev et al., 2018). Approaches to investigating this include providing images of group members containing counter-stereotypical (Dasgupta & Greenwald, 2001; Pearl et al., 2015a), and imaginary contact tasks (Blair, Ma, & Lenton, 2001). Applied directly to weight stigma, in a study on the imagined contact with a group member, participants who were provided with counter-stereotypical (i.e. "a strong, confident, attractive obese person") information self-reported less fat bias, compared with those who were provided with stereotypical information, i.e. a "weak, insecure, unattractive obese person", (Dunaev et al., 2018). Dunaev et al. (2018) note the inconsistency of the literature on reducing weight bias with counter-stereotypical information, identifying a series of studies in which no effect were found (e.g. Flint, Hudson, & Lavallee, 2013; Gapinski, Schwartz, & Brownell, 2006). However, the effect of counter-stereotypical information on reducing negative stereotypes was ostensibly supported using visual stimuli in Chapter 8, in which visual counter-stereotypical information moderately affected evaluations of larger-bodied exemplars on the competence stereotype dimension. Dunaev et al. (2018) claim that more work is needed to determine whether counter-stereotypical information affects discriminatory behaviour. Although discriminatory behaviour is not one of the outcomes of the current study, the relationship between the provision of information on exercise behaviours will be investigated. For example, in the same way that stereotypical evaluations/weight bias may be reduced through counter-stereotypical information, so too should the Proteus Effect related behavioural assimilation be affected in the current exergame study. In this study, the same stimuli used in Chapter 8 were presented to participants prior to playing an exergame. It is expected that manipulating participant's first impressions of the avatar will alter their in-game behaviours.

9.3 Hypotheses

- *H1.1* Participants in the plus-sized condition will take longer to complete the track than those in the athletic condition. The size of this effect will be of a similar magnitude to that reported by Li, Lwin, and Jung (2014).
- *H1.2* Participants in the plus-sized-running condition will take longer to complete the track than those in the athletic condition but will take less time to complete the track than those in the stationary plus-sized condition.
- *H2.1* Participants in the athletic condition, given low base rate information will over-estimate their performance.
- *H2.2* Participants in the plus-sized condition, given high base rate information will underestimate their performance.

9.4 Method

Participants

Inclusion Criteria

To be eligable for this participants were required to be over the age of 16, have normal to corrected-normal eyesight, understand written English, and be able to safely run on the spot (i.e. free from injuries). The original Proteus Effect studies had no pre-requisites that participants ought to have played videogames or identify as 'gamers', and so this was not a requirement in the current study.

Table 9.1: Age range and gender information for the participants in this study.

	Female	Male
18 - 24	41	11
25 - 34	8	3
35 - 44	2	1

Sample Size

A total of 66 participants completed the study. Of these participants, 9 were excluded because their completion time exceeded 90 seconds suggesting that they did not complete the task correctly. It became apparent that many of these cases this were due to incorrect instructions regarding the use of the motion sensing controller. Once this had been discovered, the instructions the number of participants exceeding 90s vastly reduced.

Participants were predominantly recruited from the Manchester Metropolitan University research participation pool. However, the sample was not exclusively students; some participants signed up through adverts on social media, and through word of mouth. Participants were reimbursed with a £5 Love2Shop and voucher for their time. In addition, participants recruited from the MMU research participation pool were provided with 30 participation credits.

Demographics

Demographics for the sample are presented in Table 9.1. This study had the narrowest range of participants, with the vast majority being younger women. Potential explanations for this were provided in Chapter 6, section 6.3 (page 197).

Stopping Rule

Initial analyses were run after 8 weeks of data collection. Departmental funding was acquired to collect up to 100 participants, which is comparable with the work of Peña et al, who had a study of roughly similar complexity (2X2 design with BMI as a covariate). The sample size reported by Li et al was 140. The plan nwas that if after 8 weeks (or 100 participants) there were insufficient participants to estimate the parameters with a reasonable degree of accuracy, then the study will be written up for the purpose of the thesis, but will not be submitted for publication until 140 participants have been collected.

Design

The primary study had a 2×2 between subjects design: $Performance \sim Bodytype \times Animation$. The secondary study had a 2×2 between subjects design: $Predicted\ performance \sim Body\ type \times Base - rate\ information$. This was because body type, rather than animation, was expected to have the most salient effect on predicted performance, although a three factor model will also be run and included in the model comparison analysis for exploratory purposes: $2 \times 2 \times 2$ ($Predicted\ performance \sim Body\ type \times Animation \times Base\ rate\ information$). The tertiary study had a 2×2 design: ($Stereotype\ ratings \sim Body\ type \times animation$).

Materials

The running game was developed using the Unity videogame engine. The track was designed by Tamarin Studios and purchased from the Unity Asset store. The game was controlled using the Nintendo Switch JoyCon controller which has three axis accelerometery and gyroscope input. The game was presented on a 15" Mid-2015 Apple MacBook Pro Retina Display. Two versions of the track were presented; the first had the same colour palette as provided with the track asset, and the second had a reduced number of items (e.g. stadium stands), was presented 'at night', and the colour of track was changed to grey. This 'dark stadium' was used to give the illusion that the tasks were different, when they were in fact identical. The avatars were the same white male and female models used in Chapter 7 and 8.

Figure 9.1: The main screens from the experiment, including avatar introduction scene, probability estimation, and light and dark tracks

Procedure

Participants provided their consent prior to taking part in the study. They were told that they were play-testing an exercise videogame. Participants were randomly assigned to a condition based on a CSV file that the researcher blindly randomised prior to beginning the study using the R programming language (R Core Team, 2017). Participants provided a user-name to use in the exergame. This was not stored to preserve anonymity but was intended to give participants a sense of customisation. To ensure blinding, the flow of the experiment was automated, and after the participants had been instructed on how to use the JoyCon controller and navigate the study, the researcher sat behind a screen so that the experiment was obscured.

The main scenes from the experiment are presented in figure 9.1. Participants were presented with their avatar, which was the sex that best matched their gender. The avatar was presented either running on the spot or standing idly (the same animations from Chapter 8 were used). The avatar also had either a larger body type or an athletic physique. The avatar was introduced as the name that the participant had provided. Participants were instructed to press a button to move to the next screen where they were provided with instructions. Once they pressed the button to proceed, the words “Ready, Set, Run” appeared on the screen and the participants were taken to the athletic stadium scene. As the participants ran on the spot, their avatar ran at a speed proportional to the effort they were expending. After one lap of the track, a further screen appeared asking them to take a break. On this screen they were provided with base-rate information about the second track. Half of the participants were informed that 10% of people can complete the second track in under 90 seconds. The remaining participants were informed that 90% of people can complete the second track in under 90 seconds[1]. They were asked to estimate their confidence, as a percentage, that they could finish the track in under 90 seconds. The track was actually identical to the first. After a second lap of the track, an audio cue informed that researcher that the experiment was over

Participants were then debriefed, explicitly asked not to discuss their experience of the study to maintain experimental naivety and given a £5 voucher.

Analysis Plan

A Bayesian regression analysis was run on the data using group assignment as a categorical predictor and track completion time as the dependent variable. In previous studies, the time taken to complete a course (Li et al) and the number of movements made (Peña et al) were used as dependent variables. In this study, these two measures will be highly correlated, since time taken to complete the track is entirely dependent on the number and strength of movements. Data from

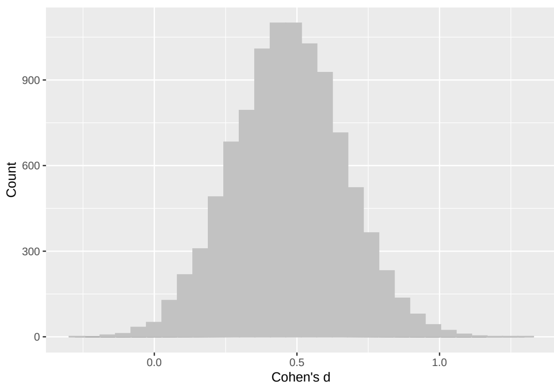
[1]This value was chosen because it is the absolute slowest that the track can be completed, if the controller was used properly (i.e. the avatar will walk around the track in this time if the controller is balanced on its side)

both DVs will be retained and explored, but the primary dependent variable will be time taken.

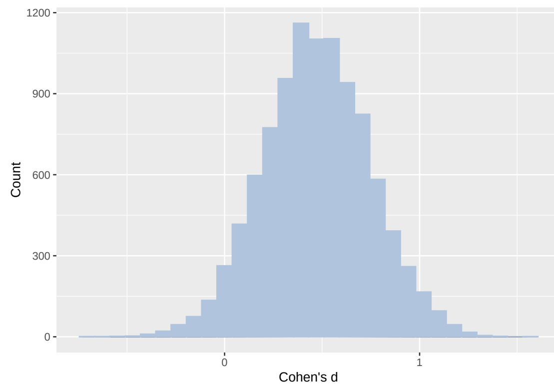
Priors

Informed priors for the initial analysis will be taken from both the previous studies from Chapter 8, and from the accumulated effect sizes from the previous research on avatar body type and exergame performance. In a meta-analysis of 46 studies into the Proteus Effect, the estimated effect size was between r = 0.22 and r = 0.26. This aggregated effect size included both attitudinal and behavioural measures, but the authors helpfully provide individual estimates for both of these outcomes. For behavioural studies, there were 27 studies, with 2396 participants, and an estimated effect size of r = 0.23, 95% CI = [0.192 0.272]. For attitudinal studies, there were 19 studies, with 1471 participants, with an estimated effect size of r = 0.26, 95% CI = [0.207, 0.310]. This is equivalent to a standardised mean difference between avatar categories of d= 0.47 and d = 0.54. The authors also present the variance for each effect size, allowing a distribution of effect sizes to be estimated. Converting the r values to d values, the distribution in Figure 9.2a may be plotted:

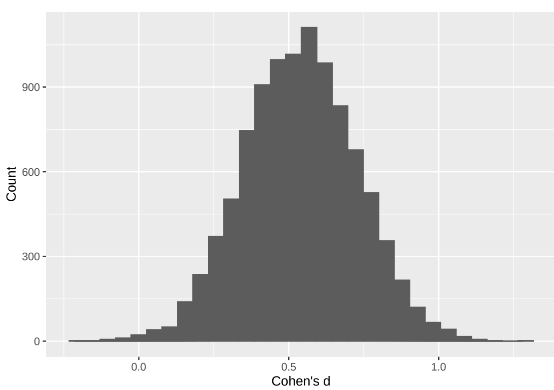
This distribution excludes the extreme low and high values reported in the meta-analysis (e.g. Kaye, Pennington, & McCann, 2018, r = 0.02; and Buisine, Guegan, Barré, Segonds, & Aoussat, 2016, r = 0.85), and there are several issues with the meta-analysis, including effect size selection which was motivated by p-values, and the inclusion of non-Proteus Effect studies in the estimate (e.g. Fox & Bailenson, 2009 did not include avatars, but virtual doppelgängers). Moreover, the effect sizes in the original Yee & Bailenson paper (2007) were potentially miscalculated, and so priors were centred on the ‘best guess’ from the meta-analysis, but was surrounded by a larger band of uncertainty by doubling the estimated variance before conversion to the standard scale, providing the distribution in figure 9.2b.



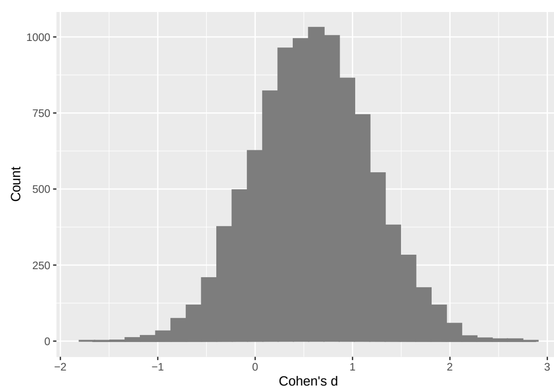
(a) Simulations taken from a distribution based on prior estimates from Ratan et al (2019)



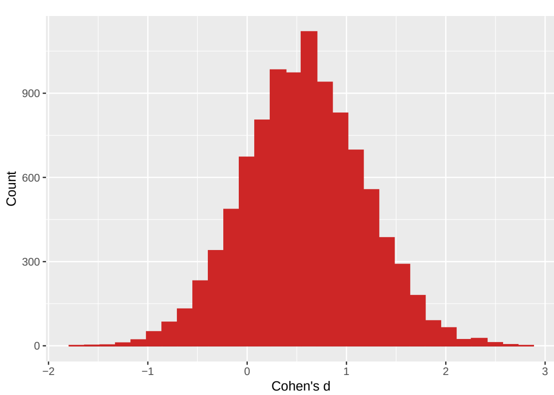
(b) Sceptical prior estimates with double the variance of Ratan et al 2019.



(c) Histogram of prior for estimates for the attitudinal Proteus Effect, taken from Ratan et al 2019.



(d) Histogram of prior distribution for generalisation, taken from McIntyre et al (2016)



(e) Histogram of prior distribution for generalisation, taken from McIntyre et al (2016)

Figure 9.2: Histograms of proposed prior distributions.

For the second part of the study, avatar body type, avatar animation, and the base rate presented will be categorical predictors, and the percentage confidence in completing the track within 90 seconds will be the continuous outcome. Since this outcome was attitudinal, informed priors were based upon the attitude effect size estimates reported by Ratan et al. (2019). Using the values provided in the meta-analysis, the distribution of effect sizes in Figure 9.2c would be expected. The variances were doubled again before being converted to d values and used as priors, with the prior distribution shown in Figure 9.2e used.

The sensitivity of the data to the priors was addressed using by re-running the models weakly informative priors used in the previous phases of the study. These were normal distributions centred on 0 with a standard deviation of 10, applied to the beta-coefficients.

Inferences

A model selection method was initially used, with predictors added incrementally until a full model was described. These included intercept only, body type, followed by the inclusion of animation. These models were compared using Leave-One-Out cross validation (LOOCV). This process was repeated for each dependent variable (performance, estimated performance, stereotypical attitudes towards larger-bodied people). Bayes Factors using the `hypothesis` function from the BRMS package in R was used to test the sign and strength of each slope and will be interpreted as continuous evidence in favour of, or against each hypothesis.

For the first hypothesis, the point of comparison will be taken from Li et al's (2014) study (r = 0.22):

1. $\beta_{body} = 0.22$.

For the other hypotheses, slopes will be tested against the hypothesised sign of the slope:

2. Including activity of the avatar as an interaction: $\beta_{act\ body} < 0$

3. Base rate neglect: $\beta_{body\ baserate} > 0$

Inclusion and Exclusion Criteria

Participants were required to have normal to corrected-normal eyesight, be able to speak English, and be able to complete the exercise (i.e. run on the spot) safely.

Participants were excluded from the analysis if they were deemed to have not completed the task correctly, i.e. not run on the spot to move the avatar. This was confirmed by the researcher who was in the same room as the participant and could hear the participants' footsteps. Participants were also excluded from the analysis if they took longer than 90 seconds to complete the task. This was the length of time it takes the avatar to complete a lap of the track when the controller is left on the table and would suggest that the participant was not running on the spot using the instructed method.

Base Rate Calculation

According to Kahneman & Tversky (1973), base-rate neglect may be calculated using Bayes Theorem 3. Participants were presented with one of two odds of completing the track in 90 seconds: 10/90% and 90/10%. This constitutes the prior likelihood of completing the course within this time. They were then asked to provide a posterior update of this likelihood given the data[2]:

$$\ni D = \exists!\ Avatar[(active \vee idle) \wedge (athletic \vee plus)] + Perceived\ Ability) \tag{9.1}$$

[2] I really wanted fancy symbols somewhere

That is, the data available to each participant consists of only one avatar, which provides them information about activity and body type; plus, their knowledge of the avatars response and their own performance on the previous track.

Given the actual base rate information that they were provided with, the likelihood ratio of success of the High Base Rate group (HBR; 9/1) versus the Low Base Rate group (LBR; 1/9) is given by:

$$\frac{P_{HBR}}{P_{LBR}} = \frac{H_{HBR} \times L}{H_{LBR} \times L} = \frac{P_{HBR}}{P_{LBR}} = \frac{9/1}{1/9} = 81 \tag{9.2}$$

This means that, absent the 'data', and were the base rate information true, a person in the high base rate condition rationally should believe they are 81 times more likely to complete the course in under 90 seconds than someone in the LBR condition.

Kahneman & Tversky (1973) plotted the median probabilities that a descriptor was judged to be an engineer under high and low prior conditions and stated that, were participants being rational in their judgements by correctly using the prior information, the plot should follow a curved line, since the probability should always be higher in the high prior group, with the likelihood maximised at 70/30. What they found was that the points fell closer to a straight diagonal (identity) line, suggesting that the prior information had little effect on evaluations.

Following this method, regardless of the condition that participants are assigned to, ratings should be substantially higher in the high prior condition than in the low prior condition. The plots presented by Kahneman & Tversky (1973) will be replicated in this study. Variance around this line will be representative of the interference that the body type and behaviour of the avatar has on attention to base-rate information.

9.5 Results

All scores were converted into z scores prior to analysis. This is because the prior distributions were constructed using standardised mean differences.

Track Completion Time

Sensitivity Analysis

The analysis was repeated with the default priors from `brms` to check for sensitivity to the prior. At N = 57, there was considerable influence of the prior selection over the posterior distribution, to the extent that the observed effect was actually reversed when possible values were constrained to the distribution implied by the Ratan et al (2019). For transparency, models are presented with both the informed and weakly informative priors. Since the weaker priors offer a more accurate description of the observed data, these will be used in the inferences, however the implications under the Ratan et al priors will also be discussed.

Model Comparison

Leave One Out Cross Validation Information Criterion comparisons for the three models under both informed and weakly informed prior conditions are presented in Tables 9.2 and 9.3.

Table 9.2: Model comparison estimates for track completion time using Leave One Out Cross Validation under weakly-informed prior conditions.

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
T1Avatar_ICP	0.00	0.00	-82.69	6.73	2.39	0.88	165.38	13.47
T1Avatar_p_wp	-1.03	0.08	-83.72	6.70	3.27	1.03	167.45	13.39
T1Avatar_Full_wp	-2.70	0.77	-85.39	6.52	4.78	1.23	170.78	13.03

Table 9.3: Model comparison estimates for track completion time using Leave One Out Cross Validation under informed prior conditions.

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
T1Avatar_ICP	0.00	0.00	-82.69	6.73	2.39	0.88	165.38	13.47
T1Avatar_P	-0.88	0.84	-83.57	6.46	2.79	0.91	167.14	12.91
T1Avatar_Act	-1.02	0.20	-83.71	6.73	3.34	1.06	167.42	13.47
T1Avatar_Full	-2.62	0.86	-85.31	6.39	4.10	1.10	170.62	12.78

The first two columns provide the relative *estimated log predictive density.* The models are ranked, with 0 being the best fit and the declining negative values indicating a poorer model fit. As can be seen, the intercept only model was found to be a superior fit for the data in this case. It is likely that this is because of lack of observations (a common theme throughout this results section). However, in the interest of completion, the full model will be explored in each of the following sections. Each section is hedged with the caveat that *more data is needed for strong inferences to be made.*

Parameter Estimates

Posterior parameter estimates for track completion time under weakly informative prior conditions are provided in Table 9.4 and informative prior conditions in 9.5. Graphs plotting these parameters, and the 95% credible intervals are also presented. As can be seen in Figure 9.3a, participants assigned to the plus-sized condition took less time to complete the track than those in the athletic condition. However, this was paired with a considerable amount of error. The posterior point estimate for the effect of avatar physique on track completion time was β_p = -0.19, 95%CI = [-0.99 – 0.59]. This estimate suggests that the majority of the probability is assigned to negative values, however a vastly non-trivial amount of probability is assigned to zero and positive value effect sizes. The ratio of evidence for the experimental hypotheses over the null hypotheses was estimated using the `hypothesis()` function in the BRMS package. The larger the evidence ratio, the less likely the data were drawn from a null distribution. Conversely, for evidence ratios closer to zero, the data were more likely to have been drawn from a null distribution. When compared with the effect size from the Li, Lwin & Jung (2014), the evidence ratio of BF_{10} = 0.06 is strong evidence against the hypothesis of the effect being at least as large as the effect from 2014. When compared against a point null hypothesis, the evidence ratio of BF_{10} = 0.45 shows that there the data were more likely to have been drawn from the null (i.e., zero

Table 9.4: Parameter estimates and 95% credible intervals for the effect of physique and animation on track completion time under weakly informative prior conditions

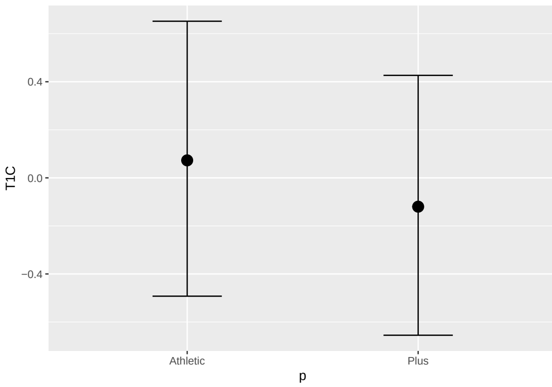
	Predictors	Estimates	CI (95%)
2	Intercept	0.07	-0.49 – 0.65
3	p: Plus	-0.19	-0.99 – 0.59
4	Act: Idle	-0.17	-1.07 – 0.71
5	pPlus.ActIdle	0.36	-0.76 – 1.51
6	Observations	57	
7	R2 Bayes	0.048	

Table 9.5: Parameter estimates and 95% credible intervals for the effect of physique and animation on track completion time under informative prior conditions

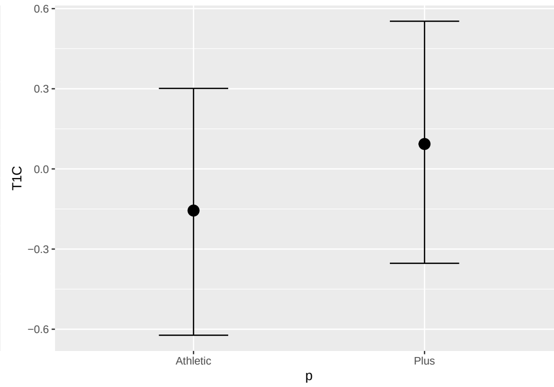
	Predictors	Estimates	CI (95%)
2	Intercept	-0.16	-0.62 – 0.30
3	p: Plus	0.25	-0.20 – 0.71
4	Act: Idle	0.06	-0.77 – 0.90
5	pPlus.ActIdle	-0.07	-1.03 – 0.86
6	Observations	57	
7	R2 Bayes	0.041	

difference) distribution than the experimental distribution - although this is only weakly supported, and does not meet the pre-registered criteria for confidence.

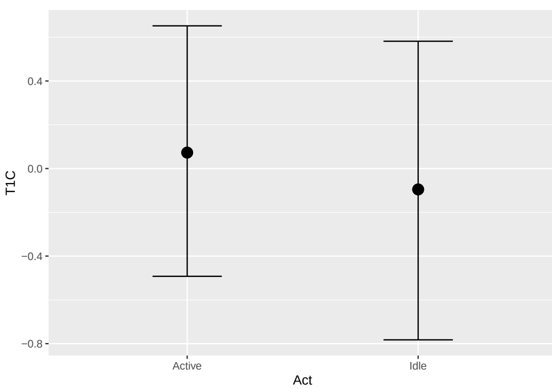
The effect of avatar animation track completion time is presented in figure 9.3cc. The point estimate suggests that those in the active condition were more likely to complete the track faster than those in the idle condition, but once again there is considerable error around these estimates. The posterior point estimate for the effect of avatar animation on track completion was $\beta_{act} = 0.36$, 95%CI = [-1.07 – 0.71]. Again, these estimates suggest that a slight majority of the probability is assigned to positive values, however a non-trivial amount of probability is assigned to zero, and negligibly small effect sizes. There was no



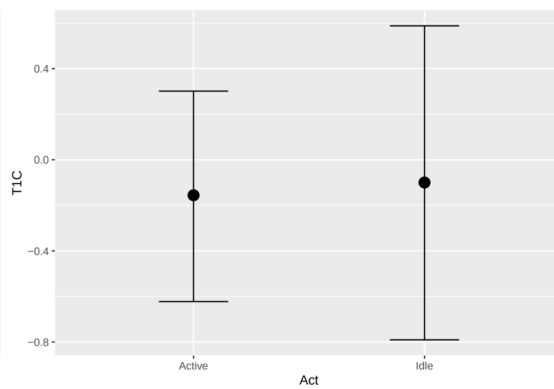
(a) The effect of physique on track completion time under weakly informative prior conditions.



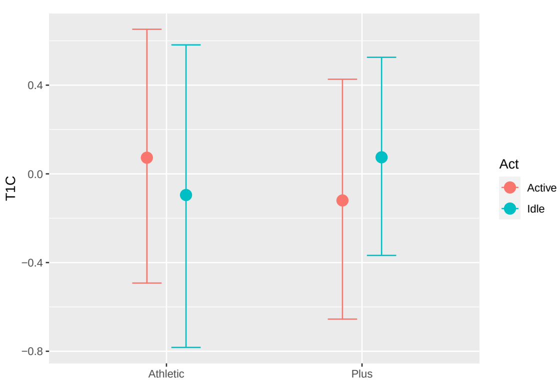
(b) The effect of physique on track completion time under informative prior conditions.



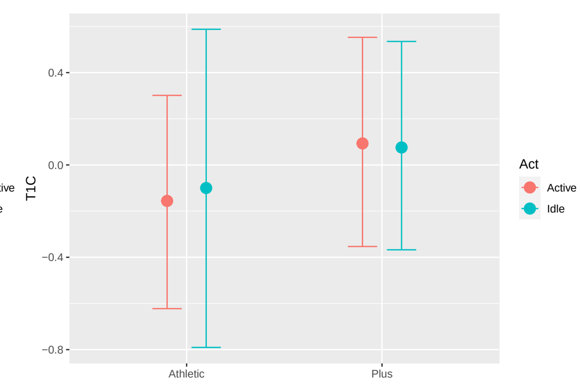
(c) The effect of animation on track completion time under weakly informative prior conditions.



(d) The effect of animation on track completion time under informative prior conditions.



(e) The interaction between physique and animation on track completion time under weakly informative prior conditions.



(f) The interaction between physique and animation on track completion time under informative prior conditions.

Figure 9.3: Plot of the posterior median estimates and 95% credible intervals of the track completion analysis (right hand panels are the distributions under informed prior conditions).

hypothesised strength of effect predicted, and so only a directional hypothesis was tested. When compared against a point null hypothesis, $BF_{10} = 0.54$ neither sufficient evidence for nor against the effect.

The interaction between physique and animation is plotted in figure 9.3d. Under the weakly informed prior conditions, the point estimates suggest that participants in the larger bodied condition completed the track faster when their avatar was running when first presented. Conversely, people in the athletic condition completed the track faster when their avatar was presented idly. However, these results must be interpreted with caution, since the distribution of posterior samples covers a range of over 1 z score. The posterior point estimate for the interaction between avatar animation and physique on track completion time was $\beta_{act:phys} = 0.36$, 95%CI = [-0.76 – 1.51]. This estimate suggests that the majority of the probability is assigned to positive values, however a non-trivial amount of probability is assigned to zero and negligibly small effect sizes. There was no hypothesised strength of interaction predicted, and so only a directional hypothesis was tested. When compared against a point null hypothesis, $BF_{10} =$ 2.83, the evidence ratio bordered on weak evidence, although further data would be needed to confirm whether this is the actual direction.

Predicted Performance

Participant's self-reported predicted performance on track two of the experiment were analysed with both the base-rate information and avatar physique as independent variables.

Sensitivity Analysis

Again, the data were strongly affected by prior conditions. This is indicative of a weak likelihood (i.e. too few observations ,or too weak a signal) and was unavoidable under the time and financial constraints. As in section 9.5, the models constructed under weakly informative prior conditions offered a more accurate description of the data and will be presented as canonical, with the same models under informative priors presented for comparison.

Model Comparison

Leave One Out Cross Validation Information Criterion comparisons for the three models under both informed and weakly informed prior conditions are presented in Tables 9.2 and 9.3.

Table 9.6: Model comparison estimates for performance estimation using Leave One Out Cross Validation under weakly informative prior conditions.

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
EstAvatar_BR	0.00	0.00	-80.97	3.42	2.35	0.26	161.94	6.83
EstAvatar_ICP	-1.16	2.08	-82.13	3.38	1.39	0.17	164.26	6.75
EstAvatar_Full	-2.03	0.59	-83.00	3.44	3.52	0.39	166.00	6.88
EstAvatar_P	-2.41	2.17	-83.38	3.45	1.89	0.22	166.75	6.89

Table 9.7: Posterior parameter estimates of the effects of base rate information and avatar physique on participants reported predicted performance under weakly informative priors.

Predictors	Estimates	CI (95%)
Intercept	0.15	-0.47 – 0.74
p: Plus	-0.69	-1.43 – 0.09
BR: BR 90	0.01	-0.80 – 0.83
pPlus.BR90	0.85	-0.19 – 1.87
Observations	57	
R2 Bayes	0.156	

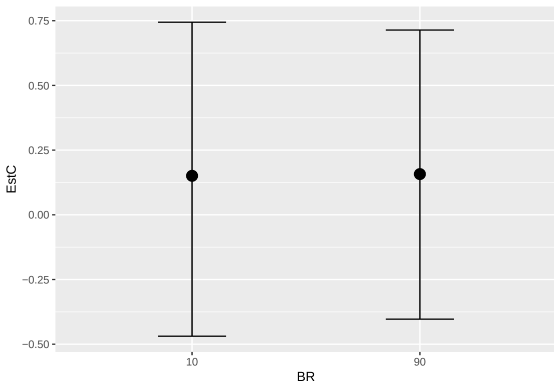
The full model ($physique \times BR$) was the best fit of the data in both cases. As such, the full model will be used throughout the following sections.

Estimates

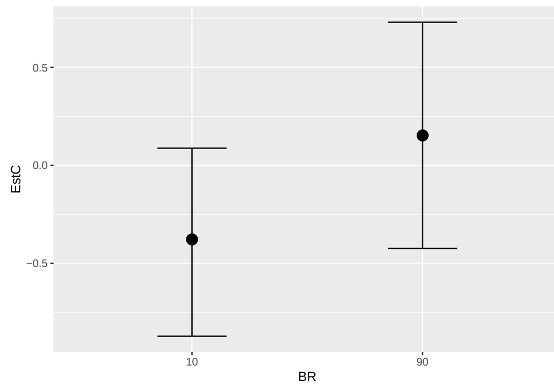
Posterior parameter estimates for participants estimated probability that they would complete second course in under 90 seconds are presented in Table 9.7 (informed priors: Table 9.8). Graphs plotting these parameters, and the 95% credible intervals are also presented in figure 9.4.

As can be seen in Table 9.7, participants assigned to the high base-rate condition were no more confident that they would complete the track in under 90 seconds than those in the low base-rate condition. The posterior point estimate for the effect of base-rate on predictions was β_{BR} = NA, 95%CI = [-0.80 – 0.83].

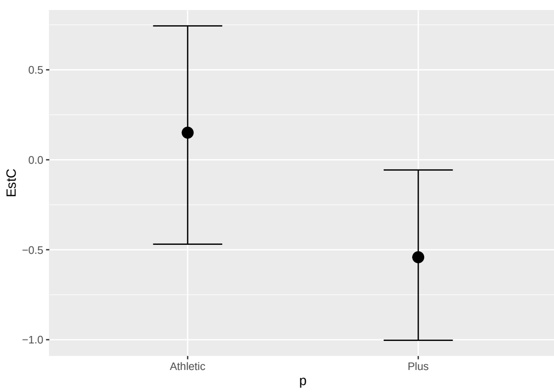
As can be seen in figure **??**, participants assigned to the athletic condition were more confident that they would complete the track in under 90 seconds than those in plus-sized condition. The posterior point estimate for the effect of physique on predictions was β_{BR} = -0.69, 95%CI = [-0.80 – 0.83]. When compared against a point null hypothesis, $BF_{10} = 6.37$ there is weak evidence that the data were drawn from a non-zero distribution.



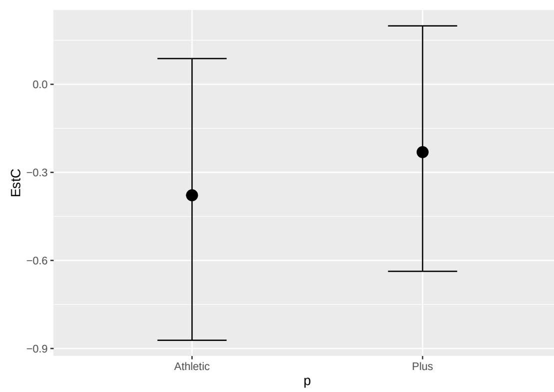
(a) The effect of base-rate data on estimated performance under weakly informative prior conditions.



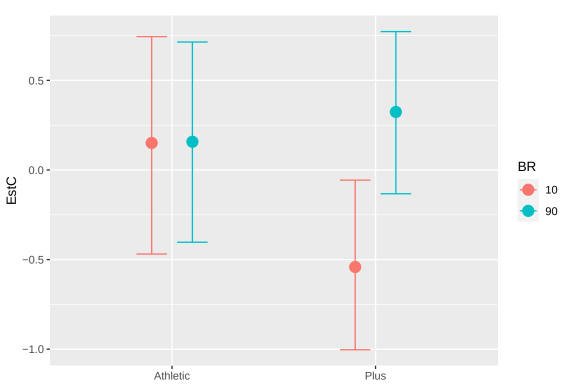
(b) The effect of base-rate data on estimated performance under informative prior conditions.



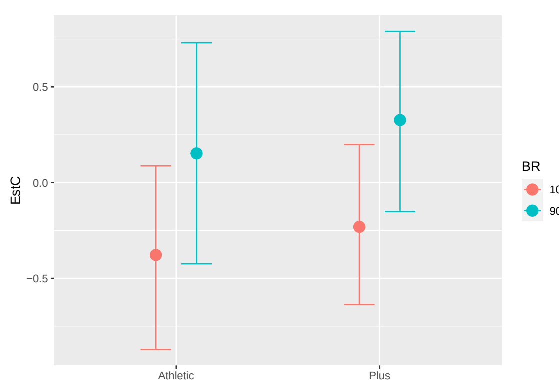
(c) The effect of physique on estimated performance under weakly informative prior conditions.



(d) The effect of physique on estimated performance under informative prior conditions.



(e) The interaction between of animation and physique on estimated performance under weakly informative prior conditions.



(f) The interaction between of animation and physique on estimated performance under informative prior conditions.

Figure 9.4: Plot of the posterior median estimates and 95% credible intervals of the track completion confidence estimate. Panels along the right hand side are the models with informed priors.

Table 9.8: Posterior parameter estimates of the effects of base rate information and avatar physique on participants reported predicted performance under informative priors.

Predictors	Estimates	CI (95%)
Intercept	-0.38	-0.87 – 0.09
p: Plus	0.15	-0.29 – 0.61
BR: BR 90	0.53	-0.20 – 1.31
pPlus.BR90	0.02	-0.87 – 0.87
Observations	57	
R2 Bayes	0.098	

Figure **??** presents the interaction between physique and base-rate. Surprisingly, participants in the plus sized condition were more confident that they would complete the track in 90 seconds than people in the athletic condition across both base-rate conditions.

Results Summary

The final data set, once exclusions were made was just over half of the planned sample size, and so caution is required when interpreting the findings from this chapter. The posterior parameter estimates under the pre-registered priors were shown the opposite trends to the raw data. This was not the case under weaker priors. This suggests that either the priors taken from the Ratan et al (2019) meta-analysis were too strong for this dataset or did not represent the actual effect. There was a small reduction in track completion time in the plus-sized condition, compared with the athletic condition under the weak priors. This is contrary to the results of the Li et al (2014) study. The majority of the distribution was negative, but the credible interval crossed zero and much of the right tail was positive. The trends in the data under weak priors suggested that participants assigned to the running-animation condition were the fastest group. But this was reversed in the informed prior condition. Under weak prior conditions, there was evidence that people assigned to the athletic condition ignored base-rate information to a greater extent that those in the plus-sized condition. This was the strongest finding in the current study.

9.6 Discussion

There were two main questions in this study: how do avatars with different attributes affect exercise behaviour and base-rate neglect. Although the results from this study begin to converge upon an answer to these questions, all must be interpreted cautiously due to the small sample size.

Interestingly, participants in the plus-sized condition appeared to run, on average, faster than those in the athletic condition. This is quite contrary to the hypotheses drawn from the Proteus Effect, as reflected in the bizarre behaviour of the models under prior conditions that favoured the Proteus Effect. In line with the meta-theory discussed in Chapter 3, particularly from page 67, this finding is vastly insufficient to contradict the Proteus Effect, and indeed the inferential methods that were pre-registered required stronger evidence against than in favour of this effect.

Feedback from one of the participants was they ran faster with the larger bodied avatar because they felt it *needed the exercise*. Because no systematic qualitative data were collected, this claim cannot be further explored, but it raises some interesting questions. For instance, it may be that when an avatar is not identified with, the exercise goals become directed at the avatar rather than the shift in self-perception required by the Proteus Effect (*they need the exercise* rather than *I am slow and lazy*). Secondly it may be that the goals of the individual may shift (*I am over weight, so I must exercise*). Both explanations are severely *ad hoc*, and it would be unfortunate for the theory, if after further testing, it because necessary for these explanations to be explored[3]. As noted, there was so much overlap in the credible intervals that the estimates could realistically be equal.

The effect of animation on track completion time was also equally noisy, with people in the idle avatar condition appearing to complete the track faster than those in the active condition. However, when the interaction between both variables on track completion time was addressed, an interesting pattern appeared to be present in the data. People in the larger bodied condition completed the track faster when they were in the active condition than when they were in the idle condition, and this was comparable to the athletic group. Once again, this needs more observations to corroborate, but if the point estimates remain stable with additional testing, this is one possible direction that the data could go.

Regarding baserate neglect, there was no change in estimates based on the high and low condition in isolation, although people in the plus-sized condition rated their predicted performance lower than those in the athletic condition. A very interesting pattern can be seen in the interaction between base-rate and physique condition. Participants in the plus-sized condition appeared to actually pay attention to the base-rate information as shown by the almost 1σ difference in ratings between the low and high conditions. The estimates of those in the athletic condition however were nearly identical, suggesting that this group ignored base-rate information based on the appearance of their avatar[4]. This observation is the strongest evidence from this study ($BF_{10} = 2.6$), although still not sufficient to draw firm conclusions from at this time.

[3]It was hoped that this study would keep the Proteus Effect on a progressive research programme, rather than make it enter a degenerating one.

[4]This was the only difference between conditions

Radical Transparency

> ...after you have not fooled yourself, it is easy not to fool other scientists
>
> Feynman, 1974

The role of the prior conditions is an essential point of discussion. The *body-related* Proteus Effect as observed by Li et al. (2014) was more likely when the prior conditions favoured their observations. With further data, it is possible that the likelihood will converge on the findings of Li et al. (2014), and the priors will become trivial. However, this demonstrates that blindly using priors that favour a theory open up a potential source of bias, particularly under small sample sizes, and reinforces the need for sensitivity checks. Indeed, by reducing the prior variance to 0.01^2, very strong evidence for the Proteus Effect may be reported[5]. Figure 9.5 demonstrates this. The left panel shows the R output for a model in which the variance of the prior was constrained, and the mean was set at the value reported by Li et al (2014). What happens is that with weak data we become virtually certain that the estimate confirms the hypothesis, when in fact the observed value was vastly uncertain and in the opposite direction[6]

Relate to the Proteus Effect

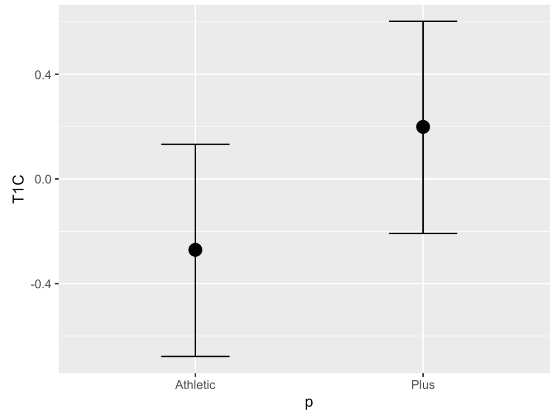
The Proteus Effect was only supported when it was assumed that the Proteus Effect existed. That is, the observed effect was morphed, much like the titular Greek god, under Proteus Effect-congruent prior conditions. If anything, this merely shows that the current dataset was not strong enough to shift the prior belief. This would be a requirement to significantly challenge, at least the

[5]Indeed, a BF of ∞ is returned.

[6]This would of course be a questionable research practice (if not outright fraud) which is why I present it in the discussion rather than the results section. But this is not unique to Bayesian statistics - I could probably do something very similar in the frequentist tradition (e.g. see Simmons et al discussed on page 70). This demonstration is simply to reinforce the importance of pre-registration in certain cases.

```
Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
Intercept       -0.27      0.21    -0.68     0.13      16767 1.00
pPlus            0.47      0.00     0.47     0.47      34142 1.00
ActIdle          0.17      0.41    -0.63     0.97       9970 1.00
pPlus:ActIdle   -0.29      0.42    -1.13     0.54      10038 1.00
```

(a) R readout showing an exact posterior estimate of the effect size that Li observed (0.47) with 0 error



(b) Marginal effect graph from this hacked model to be contrasted with the results section

Figure 9.5: Prior hacked parameters favouring the effect by Li et al (2014)

physique-related, Proteus Effect. Under the weak-prior conditions, the effect was the reverse to what Proteus Effect Theory would predict, although potential reasons for this have been discussed. It would be interesting to determine whether Self-Presence or Player-Avatar Relations as measured by the cPAX (cf page 22) moderate this effect, since the anecdotal evidence suggests that participants in the larger condition were exercising for the benefit of the avatar, rather than feeling energised themselves. This would imply an avatar-as-symbiote, or avatar-as-other type relationship. Indeed, on page 53 identification was added as a corollary to the necessary conditions, and further research would determine whether identification is a necessary condition in its own right. This could have been an interesting avenue for the current thesis, but attention was focussed on the stereotype elements of the question. One way in which the Proteus Effect theory was supported was in the finding that base-rate neglect was more likely to occur in participants using an athletic avatar, than a plus-sized avatar.

Strengths and Limitations

The study benefitted from having an entirely double blind design. The researcher was only made aware of the participants' assigned condition through remarks about the avatar (e.g. *'it's a bit fat'*), but due to the instructions given at the start of the study not to inform the researcher of their assignment, this was incredibly rare. The study also benefitted from across the board customisation; the researcher had full control over the flow of the study, the presentation of the avatars. This did amount to a fairly dull videogame (despite participants claiming that it was *fun*[7], however it isolated the act of running and the effect of the avatar responding without any other motivating factor that may be present in exergames, such as collectables, sound effects, and visual effects. This means that it is possible to build upon the design to identify other contributors to energy expenditure.

The stacked outcomes from this study were a weakness, and this was dealt with by having a primary research question which was addressed first (replicating and extending a former effect), meaning that this main effect of interest was ostensibly isolated. The subsequent research question may have been subject to more interference from the previous manipulation, meaning that less credence should be given to this finding. Indeed, these findings have raised some interesting questions which may warrant future investigation, and The Author encourages readers to engage in the field and attempt such replications. A further potential weakness was the omission of covariates such as amount of time spent gaming, or current fitness attitudes; these have been used in previous studies, but rarely (if ever) have an effect on the outcome. Checks for constructs like the experience of presence or enjoyment of the game may have provided useful information, but arguably answer a different research question and represent potentially interesting future directions.

Since the main source of recruitment for this study was the Manchester Metropolitan University participation system, the majority of the pool were women, meaning that the sample was biased (as it was in Chapter 6). Attempts will be made to balance this bias with future data collection, but the results as they currently are cannot be generalised too far.

[7]The *fun* base rate for studies in Psychology is quite low!

Applied Context

Use of athletic avatars may affect perceived competence, but there is little evidence supporting the assertion that people in the larger bodied condition performed at a lower level than those in the athletic condition. Indeed, the raw data suggested the opposite was the case. An interesting moderator might be found in the type of relationship that the participant experienced with their avatar. In the cases where participants experienced the avatar as being themselves, *avatar-as-self* relation, J. Banks & Bowman (2016a), it might be that a Proteus Effect style response is activated. However, if participants experienced the avatar as being *an other*, then this might result in a different type of response. For instance, Ahn et al. (2015) has conducted some work on the use of virtual pets to reinforce healthy behaviours such as good diet and exercise. Such an approach fosters an *avatar-as-other* or *avatar as symbiote* relation (what I called an agent, or representation in Chapter 4). Here, behaviours are performed not to benefit the self, but to benefit the digital character. This possibility was reflected in feedback given by one of the participants that they exercised because it looked like the avatar ”needed it”. Were this the case, the quote from Li et al. (2014) could be restated:

> ...giving [...] children [who identify the avatar as the self] a fixed avatar that is bigger in body size may not be a good idea if the purpose is to encourage them to exercise, [whereas it could be beneficial if they identify the avatar as an other, or a symbiote].

But ideally *using* any body type to influence exercise should be avoided.

Future Directions

The preliminary findings from this study have opened some possible avenues for the future of Proteus Effect research, health-related behaviours, and representation in general.

The focus of the five[8] studies discussed in Chapter 4 (Kuo et al., 2016; Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014; Verhulst et al., 2018) was to affect attitudes or behaviour simply through manipulating body size, in some cases leading to some questionable conclusions (e.g. removing larger bodied avatars) regarding representation in the media[9]. That is, the manipulation was reliant on negative attitudes. This study marks a move towards improving evaluations of people through stereotype manipulation: rather than the endless demonstrations of the Proteus Effect, attention might be better directed at improving stereotypical evaluations of stigmatised groups either using counter stereotypical information, or ideally balanced, diffuse, and realistic representation of groups.

Although this study focussed on exercise exertion as an outcome, a similar method could be applied to other stigmatised groups to alter impressions, and as a result behaviours. If Yang et al. (2014) suggests that playing Grand Theft Auto with a black avatar increases negative perceptions of black people and promotes aggression, then will pairing this experience with pro-social behaviour reduce this?

Moreover, this could also be applied to games like Kinect Sports Rivals from Chapter 5. Having a wider representation of body types in a game like this has the two fold benefit of allowing players to play as themselves, but also essentially having the same effect of animating people with different physiques doing athletic activities. As was shown in Chapter 8, and to some extent in the current study, such a manipulation may enhance both self-efficacy and general evaluations of larger bodied people. Indeed, even if the player is, or chooses to be thinner or more athletic, having larger bodied non-playable characters would potentially have the effect of improving evaluations of larger bodied people. This is simply conjecture at present, and would need further study, although some authors (e.g. Peña & Kim; 2014) may argue that having an "obese" opponent will reduce physical activity. Indeed, the effect with the greatest interocular impact

[8]Only the three that had a common method were discussed in this chapter.

[9]It is worth noting here that the authors of a study demonstrating that playing a game as a black avatar made white people more aggressive did not suggest that white people should not use black avatars (Yang, Gibson, Lueke, Huesmann, & Bushman, 2014).

from their study is that women were most active when both avatars "normal", with smaller differences reported in the other conditions. The Author argues that this is still not a strong enough case for the reduction of representation in videogames[10]. A longer term (multiple sessions, or multiple games) study might be interesting to determine whether such an effect (if it exists) habituates over time as opponents are evaluated as more competent. Such a study could potentially be conducted using existing videogames by simply adjusting the avatar body type and difficulty setting of the opponent AI.

As noted earlier (page 53), determining the role of self-presence, or even avatar-relation preferences (Banks & Bowman, 2013) in the Proteus Effect may help develop the theory, particularly for distinguishing between shifts in self-perception and response to the perceived *needs* of the avatar.

Findings and Conclusions

No evidence was found either for or against the behavioural (in contrast to the attitudinal) Proteus Effect in this study so far. It is likely that the sample size was insufficient to update the prior. As such, it is not possible to refute or confirm the suggestions of Li et al. (2014), Peña & Kim (2014), or Peña et al. (2016), that larger bodied avatars lead to reduced activity during exergames. Patterns in the data suggest that the $Activity|Size \times Animation$ activation observed in Chapter 8 translates into behaviour, but again it is too soon to tell whether this will form an effect upon further observations. The closest the study came to 'An Effect' was the $Arrogance|Physique$ activation, observed in the performance estimation task. Participants were more likely to ignore base-rates when they had an athletic-looking avatar, than when they had a plus sized avatar. However, the weight of evidence did not meet the minimum inference criteria that was set in the pre-registration. It is possible that with more data, this evidence ratio

[10] Indeed, this effect did not appear to replicate in the latter (Peña et al., 2016) study on men, and the authors had to resort to mining the effect using self-reported perceived difference in BMI between the two avatars.

will continue in the same direction. A potential explanation for this tentative finding comes from the Dual Process theory (Kahneman, 2011) summarised in Chapter 2 (page 56). Under this explanation, participant's estimations of their own ability were affected by the irrelevant stereotypical information from their avatar, rather than the 'factual' base-rate information that was provided. An interesting observation was the uniqueness of this observation to the athletic condition. If belief in future performance was informed by weight stigma, then the expectation would be for participants in the larger body condition to ignore base-rate information and under-estimate their ability, regardless of the 'factual' information provided. This was not the case and suggests that the mechanism between beliefs and the $Arrogance|Physique$ activation is mediated through the base-rate neglect heuristic.

This observation, if supported by future research, has applied value. Insisting that players have muscular and heroic-looking avatars, as in Kinect Sports Rivals (Chapter 5), may lead to an overestimation of their own ability. Videogames typically provide a form of 'base rate' information to players in the form of difficulty settings. Easy mode implies that the majority of first time users will be able to make progress, regardless of their experience; whereas hard mode implies that without a considerable amount of transferrable skill, or time spent on the game, most users will not progress. Providing players with elite-looking avatars may lead to the skipping of stages due to an overestimation of their ability, which in many cases, may lead to a reduction of enjoyment, frustration, and the potential for 'rage quitting'[11]. This could be particularly damaging for gamers who are experimenting with exergames as a regular means of exercise, and seems counter-productive to promoting self-efficacy, which was one of the behaviour change techniques described by Lyons & Hatkevich (2013) (Chapter 2, 18).

It would be straightforward to adapt the software used in the current study to test such a paradigm, and so this may be a future direction to take the research.

[11] Slang term for quitting a game early due to frustration

On the other hand, providing an overly athletic avatar at the earlier stages of the game where the activities are low intensity, or the user is prone to become fatigued also has negative implications, since as was shown in Chapter 7 and Chapter 8, athletic bodies are associated with competence in exercise[12]. Observing an avatar that resembles an elite athlete stumbling and getting fatigued may itself result in cognitive dissonance!

[12] A parallel would be in the game World of Warcraft, in which players can create huge battle worn characters, who could, in theory, be killed by a rat at earlier levels - a level of contradiction that may push the suspension of disbelief too far.

Chapter 10

General Discussion

This research programme adopted a mixed methods approach to investigate the influence of avatar appearance on health related behaviour (in this case, exercise). Previous research was approached critically, and this critique informed the development of the subsequent empirical studies. The project started with the question: *Can avatar appearance affect health-related behaviours?* The implications of this question were unpacked and approached from different angles and through different lenses (Willig (2013), p40). This included a thorough reading of the previous literature on the assimilation of avatar behaviours and, using a combination of Proteus Effect theory and auxiliary theories related to stereotypes and attitudes, drawing predictions of how behaviour might be changed through more complex interactions between player and avatar. The cumulation of these efforts informed a replication-extension of a previous study (Li et al., 2014). Replications are intended to test the robustness of an effect prior to developing a theory[1] (Zwaan et al., 2018), and may also be used to address the replicability of existing effects, as in various meta-scientific attempts (Ebersole et al., 2016; "Estimating the reproducibility of psychological science," 2015). It is arguably unusual to conclude a thesis with a replication, and it may make more sense to do this earlier in the programme to corroborate and expand

[1] Although until relatively recently this was rare, see discussion from page 92.

a theory before new research lines are developed. However, this was not the case in the current thesis, and the replication study potentially contributes the least information overall. It was necessary to use experimental methodology to make the connection between the attitudes research from Chapters 6 - 8, and exercise, which was ultimately the outcome of interest.

A bottom up approach was necessary to form inferences about the behaviour that were based on solid theoretical foundations, using the data from Chapters 6 to 8. That is, under the Proteus Effect (pp 30 - 40), attitudes are necessarily the antecedent to behaviour, and so a thorough investigation of the inferences made about people based on their physical appearance, and the attitudes or beliefs that these derive from was necessary before attempting to manipulate behaviour using avatar appearance. The findings from the programme will be summarised and synthesised in the following sections. Following this, strengths, limitations, reflections, and future directions will be addressed before drawing the thesis to a close.

10.1 Summary of Findings

The overall aim of the thesis was to assess whether health-related behaviours could be influenced using the appearance of an avatar. As this broad aim was refined in Chapters 4 and 5, the focus of the thesis shifted towards a critical evaluation of the Proteus Effect methodology; an investigation of potential negative consequences of embodying athletic avatars; and the development of a body-positive alternative to the recommendations made by Li et al. (2014), Peña & Kim (2014) and Peña et al. (2016). Through this line, the outcome of *health behaviours* was narrowed to *exercise*. Attempts were made to be rigorous in the design of the methods to assess the contribution of visual and behavioural information contained within an avatar, and their effects on energy expenditure during an exergame.

Chapter 4 highlighted that based on their findings, Li et al. (2014), Peña &

Kim (2014) and Peña et al. (2016) suggested that larger bodied people should be restricted to ostensibly athletic avatars within the "normal" BMI range; however the accounts of the gamers in Chapter 5 suggested that while athletic bodies were not stigmatised, they appeared not to be as revered as intended. Moreover, the avatar size restriction recommended by previous research could be interpreted as merely reinforcing objectification and unrealistic norms. Chapters 6 and 7 showed that larger bodies do induce stigmatising evaluations; and yet the results from Chapter 8 suggest that the provision of counter-stereotypical information may reduce stigmatised ratings. In addition, it was suggested in Chapter 5, and broadly supported in Chapters 6 to 8 that stereotypes related to both the stigmatised *larger bodied* group, and the reportedly revered *athletic group* are actually ambivalent, meaning that positive information may be drawn from the stigmatised, and negative from the revered. Finally, in Chapter 9, preliminary support was presented for a minor alleviation of the negative effect of body size on exergame exertion discussed in Chapter 4 when players first saw their avatars presented running, compared with standing idly.

In line with the mixed design approach (Johnson & Onwuegbuzie, 2004), the programme was executed in different phases, each addressing a different objective. These will be summarised below.

Exploratory Phase

Chapter 4

The first objective, addressed in Chapter 4, was to assess the strength and quality of evidence that avatar appearance may affect health-related behaviours (i.e. *what is already known?*).

The systematic review highlighted that there is little research, and much heterogeneity in the approaches to avatar-based health research. Indeed, only three studies, each on three different groups (children, men, and women) similar enough to compare directly (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014). Two further studies had similar independent variables (manipulating avatar body size), but different dependent variables and found little to no effect (Joo & Kim, 2017; Verhulst et al., 2018). The three studies that found an effect were subsequently used as the foundation for the empirical portion of the thesis. The authors of the studies suggest that the Proteus Effect was the cause of the differences, and therefore larger bodied avatars invoke sufficiently salient stereotypes (e.g. laziness), leading to participants assimilating behaviours related to these stereotypes (i.e. slower activity) into their own self-schema. It was the suggestion that larger bodied people (specifically children in this case) be restricted to thinner avatars by Li et al. (2014), and the participant's accounts from Chapter 5 suggesting this would be unwanted, that provoked the critical line of research in the thesis. This line involved challenging the suggestion of Li et al. (2014) through an in-depth exploration of the stereotype structure associated with larger-bodied and athletic individuals, and in addressing the methodological weaknesses of the previous studies.

Chapter 5

The next objective (addressed in Chapter 5) was to explore the experiences of creating avatars, and of being automatically portrayed as an idealised self-similar avatar by a videogame. Although larger-bodied people were not purposively sampled[2], Kinect Sports Rivals is an example of a game in which the recommendations of Li et al. (2014) are taken: players do not have access to larger or untoned bodies. In this qualitative study, participants discussed their approaches to avatar design which included: making minor changes to their self-similar avatars, representing their ideal selves whilst keeping their identity intact; creating entirely different characters that also represented their ideal

[2]This study was conducted prior to the refinement of the aims to address weight stigma. It may also be difficult to recruit from this group because of difficulty in finding non-stigmatising language to advertise such studies.

selves; role playing people and creatures they could not *be* in the real world. Some participants discussed not wishing to play with *fat* or *ugly* avatars[3]. A critique of the limitations of body types available in *Kinect Sports Rivals* was inferred from the accounts of some participants, but others expressed relief at not being portrayed as fat. The evidence of *fat phobia*, as well as the critique of *overly athletic* bodies led to the focus of the research programme from Chapter 6 onwards, since it seemed that avatars with neither body type were particularly desirable.

Chapter 6

The Proteus Effect is the dominant theory of behavioural change through avatar appearance. It requires that a salient stereotype be associated with an avatar in order for a behaviour to be assimilated. A third objective, developed throughout Chapters 6 - 8, was to investigate stereotypes invoked by larger and athletic body types. Larger bodies were indeed associated with negative *out there* stereotypes, and athletic bodies positive ones. Three factors were identified which broadly mapped onto the Warmth and Competence dimensions of the Stereotype Content Model (Fiske et al., 2002). These were:

- Determination
- Arrogance
- Affability

Each group was evaluated with positive and negative (i.e. ambivalent) attributes associated. Broad support for these constructs, and variation in the activation of these constructs between avatar physiques was provided and informed the latter confirmatory phase of the programme.

[3]Whether this represented a *feared* or *undesired* self (Higgins, 1996; Markus & Nurius, 1986; Ogilvie, 1987) was unclear, but there was a discussion of aesthetic enjoyment: why would they want to play with such avatars, when they can play with thin attractive avatars?

Confirmatory Phase

Chapters 7 & 8

Inn Chapters 7 & 8 conceptual replications of the structure found in Chapter 6 using images and videos of exemplars as stimuli were conducted. Across all modalities, athletic individuals were seen as competent but arrogant, and plus-sized people were seen as incompetent but friendly/not arrogant.

The factor structure, and relations between physiques were weaker in the visual condition (although this was difficult to quantify), suggesting that there was some interference during instantiation. This may be due to physical features such as facial expressions were included in the evaluation Or there may be a disconnect between internalisations of groups and the exemplar. Media depictions have been shown to vary between genders, for example athletic-looking women are often portrayed as thinner rather than muscular in social media (Talbot et al., 2017).

A secondary objective regarding whether the evaluations from Chapters 6 and 7 could be manipulated using counter-stereotypical behavioural information was addressed in Chapter 8. The exemplars were presented with animations portraying idle or active behaviour. This line of inquiry was informed by previous research on the effect of counter-stereotypical information on evaluations of larger bodied people which had some success in imaginary contact tasks (Dunaev et al., 2018), but an unclear effect using images (Pearl et al., 2015a) (see page 231).

In Chapter 7 (page 256), an ontological approach to stereotypes was discussed, whereby the evaluation of an instance of a group inherits attributes, categories, and relationships from the higher order class (archetype/prototype) of that category. The prototypical approach to stereotypes involves asking participants what attributes a *typical* member of a group has (Stephan et al., 1993) (see page 184). A similar approach to this was used in Chapter 6, but participants were asked to evaluate a group, not a single individual, and the response variable was an *out there* (second order, belief about the beliefs of others), rather than an *in here* (first order, subjective belief) stereotype construct. The similarity in evaluations between Chapter 6 and 7 suggests that the models were successful instances of the archetype. However, it was also briefly noted that the inheritance of attributes was not lossless, as observed by a diminished effect of physique on evaluations in the visual condition. This observation was supplemented with a footnote directing readers to an appendix of word clouds which suggest that there may be more words produced in response to the verbal descriptors than the visual exemplars (page 257) .

Based on this inheritance theory, it was hypothesised that the negative attributes inherited from the prototypical representation could be diminished using counter-stereotypical information[4]. Although this framing is consistent with the pre-registered hypothesis, further work is needed to develop this theory.

Chapter 9

The final objective was to determine whether behaviours that were consistent with the stereotypes found in the previous chapters were observed in a group of participants when they used the virtual exemplars as avatars.

As far as reasonably applicable, attempts were made to follow the necessary conditions identified on page 52, since the studies described in Chapter 4 that did not find evidence of avatar physique on behaviour (Joo & Kim, 2017; Verhulst et al., 2018) strayed from these conditions, either through no immediate opportunity to perform a behaviour, or insufficient control of the avatar.

It was predicted that if the Proteus Effect was evoked: *a.* people using athletic avatars ought to be more arrogant; *b.* people using larger bodied avatars ought to be slower; and *c.* those who used larger bodied avatars that were animated ought to be faster than the stationary larger bodied group, but slower than the athletic group.

[4] I do not say negated, because the effect was not that large (i.e. the probability of negative competence evaluation did not drop to 0).

The data collected in Chapter 9 are not yet strong enough to corroborate (or refute) the hypothesis. Moreover, the results were heavily dependent on the prior. When the prior information from the previous studies was used to inform the posterior distribution, there was weak support for the original study; however, when the priors were relaxed, the effect seemed to flip - with people in the larger body condition completing the track faster than those in the athletic condition. In the actual observed data, the latter was the case, i.e. the people in the larger bodied avatar condition completed the track faster (Tautologically, the Proteus Effect is most likely in a world where the Proteus Effect is likely)[5].

Although the action of the avatar upon first exposure was not a strong predictor of track completion time, the interaction between physique and action under a weak prior distribution suggested that people in the larger body condition completed the track faster when the avatar was running on the spot. Again, more data will be necessary for these estimates to become more stable and informative.

Regarding the effect of base-rate information and physique on predicted performance, this was again heavily dependent upon the prior specification. Under the prior conditions for the *attitudinal* Proteus Effect taken from the Ratan et al. (2019) meta-analysis, people in the low base-rate condition were less confident in their ability than those in the high base-rate condition, and there was little effect of avatar physique. However, when the range of possible values was expanded, it appeared that people in the athletic condition did not pay attention to the base-rate information, i.e. the estimates were the same for both values. However, for people in the plus-sized condition, base rates appeared to be attended to. Estimates were lower (by nearly a whole standard deviation) when participants were in the low BR condition, than those in the high BR condition. That is, there was evidence of base-rate neglect in the athletic condition, but not the plus-sized condition.

[5] This is a confusing feature of Bayesian statistics, and actually one of the strengths of the approach. It means that stronger evidence is required to dethrone a theory, because belief in the latter is stronger. However, for this reason (i.e. that it can lead to weird inferences), it is important to run checks to ensure that the researcher is not fooling themselves.

10.2 Synthesis of Findings

Weight Stigma

Weight stigma, or fat phobia, has been researched for decades (Bacon et al., 2001; Robinson et al., 1993). Research on the evaluation of larger bodied people started as early as 1983, focussing on negative attributes and stigmatisation of this group. Much later, positive evaluations of larger bodied women associated with warmth were reported (De Caroli & Sagone, 2013). Positive attitudes towards people with larger bodies were present in some of the accounts in the qualitative study of Chapter 5, with one participant (Meryl) discussing the *Find Dancing Man* internet meme (Figure 10.1), in which a larger bodied man was *fat shamed* for dancing at a concert. He was subsequently found on Twitter and invited to various dancing parties as a body-positive role model.

(a) The original post on a viral website.

(b) The result of the findDancingMan campaign on Twitter

(c) One of the Dancing Man parties

Figure 10.1: The Find Dancing Man Meme

Interestingly, according to the findings in this thesis, friendliness and outgoing-behaviours may form part of the larger-bodied stereotype structure as found in Chapters 6 to 8. It might be that this element of the stereotype is helpful in future work on positive representations of larger-bodied people. It is important to note that even though this could be seen as a positive attribute, it is still a stereotype, and encourages the expectation of larger bodied people to being happy, bubbly, and sociable, instead of making judgements based on the individual. Further, it is unclear whether the intention behind these evaluations is positive: someone can be seen as outwardly friendly, but still disliked (see Negative Valence Asymmetry from page 267). A qualitative study unpacking seemingly positive stereotypes would help elucidate this. It is also worth noting that given the Negative Valence Asymmetry model, warmth and morality are harder gained and more easily lost - and so even if positive inferences about warmth are made about a larger bodied individual, it would take only a few negative experiences to ruin this "advantage" (Skowronski & Carlston, 1989, 1992).

Contrary to some of the findings reported in Pearl et al. (2015b), counter-stereotypical information appeared to reduce the negative evaluations of larger bodied exemplars. Pearl et al. (2015b) found that neutral (neither stigmatised nor active) depictions of larger-bodied people, rather than depictions of the counter-stereotype of exercise, were associated with a reduction in fat-phobic ratings. This may be because in Pearl et al. (2015b) only images were used, and these were just of women. Static depictions of exercise may be insufficient to portray an individual as active[6], whereas being animated is more informative in terms of AST. That is, the evaluation is built from two systems (Visual and Behavioural) rather than just one.

The dependent variables were also different between the two studies. In

[6] As a case in point, Instagram gym-selfies: these give a sense of location, but not activity. The audience is expected to infer from this that the model is active. In fact, the only behavioural evidence provided is that the model has taken a selfie!

Pearl et al. (2015b), participants evaluated the group "fat people", whereas in Chapter 8 participants evaluated a single person. The former requires individual to group generalisation (McIntyre et al., 2016), i.e. an upstream influence of individuals on the group[7]. This is generally stronger for negative information than positive information in stigmatised groups (Paolini & McIntyre, 2019).

Although more work may be needed to improve evaluations of individuals before attempting to change group-level evaluations, there does appear to be a shift further up the hierarchy at a societal/media level. For example, there has been a relatively recent shift towards empowered larger bodied celebrities in the mass media. For example, musicians such as Beth Ditto from The Gossip, and more recently Lizzo (Figure 10.2d). The latter presents both themselves and their supporting dancers in stark contrast to representations of people with similar physiques from the past three decades. In music videos such as Olivia Newton John's *(*Physical) from the 1980's (Figure 10.2a), *Vindaloo* by Fat Les in the 1990's (Figure 10.2b), and *Lazy* by Xpress 2 in the early 2000's (Figure 10.2c), larger bodied people are depicted as objects of ridicule, lazy, and unfit. Compare these representations with the energetic, confident, and active personalities of the 2010's, these personalities are the very opposite of the *fat phobic* stereotype, and certainly a welcome departure. The current set of studies, with regards to the positive elements of person evaluations, could be seen as evidence for a shift in attitudes towards different body types[8]. Although there was still evidence of stigma, this could be manipulated with behavioural information, suggesting that the general public may be more open to the possibility that larger people can be energetic.

[7]In the ontological language discussed in Chapter 7, this could be seen as the instance affecting. In this regard, such a metaphor may be slightly impoverished since it would not necessarily allow for such recursion.

[8]Thanks to Sydney Y Wood for directing me to this contrast at Meta-Science 2019

(a) Get Physical by Olivia Newton John (1980s).

(b) Vindaloo by Fat Les (1990s)

(c) Lazy by XPress 2 (2000s)

(d) Lizzo, (2010s)

Figure 10.2: Depictions of larger bodied people over the past three decades.

Athletic Stereotypes

A novel part of this thesis was the provisional evidence of a salient stereotype associated with the group of people who have athletic-bodies, and that this may be also be invoked using a synthetic digital exemplar. As discussed in Chapter 2, research on stereotypes associated with athleticism typically focussed on the “dumb jock”, or “natural black athlete” stereotypes (Diekfuss, 2012; Sailes, 1993; Wininger & White, 2008, 2015). In Chapter 6 however, intelligence did not form a strong part of the idiographic element of the construct development and was subsequently not included in the list of words used throughout. In hindsight it would have been beneficial to include these words artificially to see if there is, in fact, a distinction between athletic and *athlete* + *athlete* + *student* groups. Regarding the “Naturally Gifted” stereotype, there was an abundance of words such as *dedication, conscientiousness, and commitment* that are not necessarily congruent with this stereotype. In the literature this appears to be a race specific representation, and these words were derived from the simple descriptors "People with athletic or fit body types". An interesting direction would be to use race-specific descriptors, e.g. "Black people with athletic or fit body types" if this was to be explored further. It would be useful to determine whether the "natural-black athlete" stereotype is reflected in the word generation and rating, and whether negative evaluations may be affected in a similar way to the effects reported in this thesis. Further, the same set of words could be tested with black exemplars to see if evaluations differ between races.

The stereotype that emerged appeared to be one of competence but arrogance[9]. The unique contribution here is that the stereotypes were drawn specifically from body type, with no other information about the group or individual provided. Under these conditions, intelligence was not a major defining feature. It may be that when *athlete* is conjoined with *college/university*, that a more complex process occurs in which the probability of the target getting into university for academic merit are weighed against admittance on a sports scholarship[10], leading to the assumption that the target is athletic but unintelligent. This leads to the question of if and how a target may be presented as both intelligent and athletic. Now that a baseline has been set for how body types invoke evaluations, these could be built upon by providing additional information about the model (e.g. "this exemplar is a student").

The evidence should be approached critically since the ideographic part of this work was only based on the word production of 56 people (28 per modality). Casting a wider net around the attribute space using more samples may have captured other constructs such as intelligence, and collecting demographic data would have allowed cross-cultural differences, which may have been missed in the current thesis, to be identified.

[9] It might be that the "dumb jock" stereotype is more associated with North American than European Universities, although there are insufficient grounds to base this on in the current thesis.

[10] These are less common in the UK.

Connecting Attitudes to Behaviours

The manipulations of attitudes did not reliably translate into the expected behaviours in Chapter 9. The strength of evidence in either direction was insufficient to either corroborate or contradict the Proteus Effect. Given the strength of the effect of exemplar physique on competence observed in Chapters 7 and 8, it would be expected that the Proteus Effect would be optimised: i.e. of all possible candidates for a Proteus Effect study, body size is amongst the most evocative of stereotypes. However, if anything the opposite trend was observed, with larger bodied avatars eliciting the faster track times. Participants obviously had no way of knowing that the alternative was an athletic avatar; nor would someone in the athletic condition necessarily feel that they should move slower because the avatar did not seem it needed exercise.

There are various possible reasons for this failure to replicate, the most obvious being the small sample size. One possible explanation is a lack of identification with the avatar. For instance, one participant stated that they ran fast because they felt the avatar needed more exercise because of its (larger) body size[11]. As stated, there were no measures of self-presence (briefly discussed on page 25), identification, or even immersion, so this cannot be explored further. The game in Chapter 9 was presented on a laptop screen, the controls were not the most responsive that they could have been, and the avatar was presented from a third person perspective. As such, presence may not have been particularly high. According to Galvan Debarba, Molla, Herbelin, & Boulic (2015), perspective does not necessarily affect the embodiment of avatars. A hypothesis could be drawn that under conditions of high presence, participants would assimilate *negative* behaviours into their self-schema, whereas under conditions of low presence, the participant treats the avatar as an *other* rather than as themselves and is motivated differently. Such a formulation could be seen as degenerative however (see page 67), since it is an attempt to patch up

[11] This was presented as a corollary to the necessary conditions on page 52.

an existing theory based on failed replications. None of the previous studies measured or controlled for presence, and so this would raise the question as to why it matters now, and not in the studies by Peña et al and Li et al.

Anecdotally, the majority of participants in the Chapter 9 study reported that they found the task tiring and enjoyable, further suggesting that exergames may be a viable alternative to traditional exercise, and one that may reduce perceived barriers to health discussed from page 12. If exergames are targeted at people who are in the pre-contemplation phase of health behaviour change (Prochaska & DiClemente (1983), see page 13) this may contribute towards forming of habits. Moreover, from the discussion of body-type representation in Chapter 9 (from page 332), playing exercise videogames with a wider representation of bodies may help shift perceived social norms that are included in the predictive model of the Theory of Planned Behaviour (Ajzen (1991), see page 18) as well as increasing perceived behavioural control (i.e. self-efficacy) in people who have larger, or even small and less athletic, bodies.

General

This project offers further corroboration to some of the theoretical foundations discussed in Chapters 2 and 3. For instance, the proposed data structure described by Associated Systems Theory (Carlston, 1994) was corroborated to some extent, since it had some success in making predictions for the provision of behavioural information on evaluations in Chapter 8. Chapters 6 through 9 investigated the effect of adding progressively more information to an exemplar on evaluations. In terms of Associated Systems Theory, this relates to the Visual and Action systems. There was a non-trivial effect on the word evaluations when exemplars were depicted running on the spot. In addition to these systems, AST includes an Affective and a Verbal system (see Figure 2.6, page 42). According to AST, visual and behavioural information will activate verbal and affective systems when information from these systems is scarce (Carlston, 1994). This is reflected in the data to some extent. Affective information appeared to be elicited based on words such as *happy and moody.* Moreover, elements from the Big Five personality construct (Digman, 1990)[12] were represented in the factor structure of the stereotype studies. For instance, evaluations of higher Openness, Extraversion, and Agreeableness in larger bodied people was implied by the word-applicability ratings of *fun, funny, kind*, whereas higher conscientiousness and lower agreeableness and openness was implied by the word ratings of *committed, dedicated, boring, and aggressive* in response to the athletic people. Although no attempts were made to adjust *morality or warmth* (arguably the base-rate neglect section of Chapter 9 captures some of this construct), the large effect of counter stereotypical information on competence evaluations lends partial (one-sided) support to the Negative Valence Asymmetry hypothesis (Skowronski & Carlston, 1987, 1989). Further manipulations of warmth/morality need to be attempted before firm conclusions may be made.

10.3 The Implications

The work presented in this thesis suggests that positive portrayal of larger bodied people may be possible using counter-stereotypical information. The findings may be useful when applied to both developing behaviour change interventions, and in the development of games and advertising campaigns.

Advertising and Games Development

Previous research has suggested that adverts portraying ultra-thin models may negatively affect self-perception (Halliwell, 2013), whereas "average" (on the BMI scale) models may positively affect body image in certain groups[13]. That is not to say that athletic people ought to be excluded from advertising campaigns, but they

[12] OCEAN: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism.
[13] e.g. Those who internalise cultural ideals.

should not be exclusively used[14]. Indeed, in some cases viewing athletic bodies can be motivating (e.g. Halliwell et al., 2007; Pridgeon & Grogan, 2012). However, the exclusive depiction of happy, thin, athletic people on gym advertisements may carry rhetorical information about the type of bodies that are welcome and presenting equally competent people with a variety of body types might have a positive effect on the persuasive features of gym adverts (Barlösius & Philipps, 2015). Further, as discussed in Chapters 6, 7, and 8, and particularly from page 259, such depictions may elicit feelings of malign envy, if the audience perceive the models as displaying hubristic pride.

There are also implications for games development. Previously, authors have warned about limiting the diversity of avatar affordances in videogames, both in terms of availability and restricting avatars to stereotyped representations (Kafai et al., 2007a; Kolko, 1999; Nakamura, 2013). At the same time, authors have suggested that bodies ought to be limited, particularly in exergaming contexts, for some groups to maximise outcomes (Li et al., 2014; Peña et al., 2016; Peña & Kim, 2014). Discussion of why this would not be advantageous and potentially harmful (if by harm we mean preventing people who, but for exergames with representative avatars, would not exercise) is presented in Chapter 5, in particular from page 177. Between the myriad goals and preferences that people have when designing their avatars, and with the findings that negative evaluations of larger bodied avatars may, in part, be alleviated from Chapter 8, efforts should be made to improve evaluations of stigmatised groups through positive representations, rather than excluding them altogether.

Policy and Behaviour Change

A good deal of the intervention papers that were reviewed throughout this thesis are framed in terms of *battling an obesity epidemic*, with this being cited as a

[14] This needs reiterating, I am NOT saying that athletic people should be excluded from advertising.

leading cause of death[15]. Arguably, the defining feature of obesity is appearance, and some authors have argued that focussing intervention efforts on a population based on their appearance, particularly with such negative portrayals and the prevalence of weight stigma, may be ineffective, and focus should be placed on body functionality, rather than appearance (Alleva et al., 2018; Alleva, Martijn, Van Breukelen, Jansen, & Karos, 2015; Mulgrew et al., 2018).

If running on the spot is to be considered to exemplify functionality, then there is a degree of support for this assertion in the results of Chapter 8. The combination of stereotype content and negative valence asymmetry used in this thesis may be applied to health-related intervention tailoring and advertising. This combination predicts that increasing the perceived competence of a larger bodied person may be easier than increasing perceived warmth of athletic people. This is supported somewhat by the results in this thesis.

Given the general effect of activity on evaluations observed in Chapter 8, it could be that the addition of activity to a range of body types, either in health promotion materials and adverts, or in videogames, could be used to tailor behaviour change interventions aimed at increasing healthier behaviours (e.g. increasing physical activity, or reducing sedentary behaviour). This may call for a shift in framing from tackling an *obesity epidemic*, to an opportunity to increase physical wellbeing and functionality across all body types. That is to say, the shape and size of a person's body, and their self-perception that one is unfit are not mutually exclusive[16], and actual fitness and body shape are reportedly orthogonal (Lee et al., 1999). A future line of research might present an active exemplar that is tailored to inactive or sedentary people with smaller physiques.

[15] To name a few: Jin 2009, p765; Lwin & Malik (2012); Song, Peng & Lee (2011); Kuo, Lee & Chiou (2016).

[16] As an anecdotal example, people with all types of body claimed that they were *really unfit* prior to participating in the exergame study in Chapter 9.

10.4 Critical Evaluation

Strengths

A strength of the work presented in this thesis is the use of mixed methodology, which allowed for a more nuanced approach to the question. The variety of influences on the use of avatars is non-trivial, and evaluation of an avatars by a gamer involves myriad socio-cognitive processes. For instance, how people identify, and interact with their avatars will depend on both social-psychological (e.g. self-perception) and sociological factors (e.g. the influence of media on conformity to a set of body-types). This is to name but a few, and many were not even discussed in the thesis. For example: sensory (immersion in a virtual environment), perceptual (the experience of presence, whereby *reality* becomes futile) factors may contribute to the experience of avatar use. These factors were discussed in the interview study but were more associated with poor controls and game design[17]. For example, some participants mentioned a disconnect between their movements and the avatar. The use of qualitative methods allowed these influences to not only be identified in accounts surrounding the topic, but also for them to be recognised in secondary sources also. That is, The Author could interpret the articles as texts, and infer and challenge the values expressed.

Although there were some weaknesses, the methods used in the quantitative studies were rigorous and attempts were made to reduce bias. The avatars used in Chapters 6 to 9 were designed in collaboration with Body Image experts and involved three revisions. The methods for selecting the words is entirely reproducible, and materials are provided for this method to be replicated also. There is also transparency about the weaknesses of the study throughout, and action taken to approach these weaknesses. For instance, a replication of the first study is currently underway, with 180 participants having taken part at the time of writing. The Author was also critical of his own decisions regarding

[17] These are arguably beyond the scope of a Psychology PhD, and more in line with Human Factors or Ergonomics.

priors when running the Bayesian analyses, and used a selection of alternatives throughout. The development of a bespoke experiment using Unity was also a strength of the current study. This allowed full control over the assignment, blinding, and measures of the study. Few previous studies have made such considerations over the design of the experiment.

With the exception of Chapters 5 and 6[18], all studies in this programme were pre-registered. This means that the methods, hypotheses, and analysis plans are stored on a read-only repository for reviewers and audiences to refer. A consequence of this is that none of the hypotheses can be presented as confirmatory if they were formed after the data were seen which was a common practice until recently (John, Loewenstein, & Prelec, 2012).

The programme was founded on a rigorous systematic review. All available research in this area was reviewed, and the weaknesses in the methods that were identified were addressed in the quantitative studies.

Finally, the appropriate analysis methods for each study were researched prior to analysis in the majority of cases. The programme research question was addressed in terms of the confidence that should be placed in the Proteus Effect, given previous studies and new knowledge generated through empirical study. Rather than a frequentist approach, which would require repeated testing and inferences drawn from a long run probability, a Bayesian approach that focusses on updating existing knowledge was deemed to be more appropriate for answering to this question.

Limitations

Missing Moderators

In planning the studies in this thesis, priority was given to being economic and reducing the number of parameters that would need to be estimated, and the cost

[18]The replication of 6 has been pre-registered.

of running each study. Indeed, adding more surveys to Chapter 7 or 8 would have had the twofold issue of requiring more parameters to be estimated, and increasing the time taken and therefore reducing the number of participants that could be afforded. In the following paragraph, the various moderators that may have increased the value of this thesis will be discussed.

Body Size, Perceived Controllability, Cognitive Dissonance As noted in conversations with colleagues, questions at conferences, social conversations, I included no measure of body-related *group membership*. That is, I asked no-one about body size; how people evaluated their bodies (body satisfaction); or take any objective measures of adiposity. There exist few reliable and valid methods of ascertaining adiposity. The most widely used measure is Body Mass Index ($\frac{Mass}{Height}^2$), however there is evidence that this is not a valid measurement, and provides little information about body composition (Grogan, 2017; Chapter 2). From discussions with physiologists[19], the so-called *Pinch Test* is a reliable and valid measure of adiposity; however the majority of participants in the quantitative side of this study completed the study online, so a pinch test was not possible! Moreover, discussion of body size may have further primed the purpose of the experiment. A further reason for not taking a body composition measure was that weight stigma appears to be independent of group membership (Pearl & Puhl, 2018; Puhl & Heuer, 2010; Täuber et al., 2018), although the *level* of implicit negative attitudes has been shown to be negatively related to body-type, i.e. thinner people have stronger implicit anti-fat attitudes (Schwartz, Vartanian, Nosek, & Brownell, 2006).

A possible moderator for the effects in this study is perceived controllability of body size. It might be that negative evaluations of larger bodied people are stronger in those who believe that body size is entirely within the persons control, and it is through poor choices surrounding activity and eating that a

[19] I thank Aaron Caldwell and Megan Rosa-Caldwell for and interesting discussion surrounding this at SIPS 2019.

person has a larger body. In support of this possibility, Tiggemann & Anesbury (2000) demonstrated that there was a weak-moderate correlation between negative stereotypes of fat people and belief in the controllability of body size in children[20]. However, in a subsequent study, an intervention to change perceived weight controllability in another group of children was not associated with reduced negative stereotyping of larger bodied people, despite perceived controllability being reduced (Anesbury & Tiggemann, 2000). Controlling for perceived controllability would have been an interesting additional moderator in this thesis, however, the findings from Chapter 8 lend support to the assertion that perceived effort also a strong moderator (we cannot compare this with controllability). Black, Sokol, & Vartanian (2014) found that vignettes providing a depiction of a larger bodied person who was trying to lose weight reduced negative stereotyping to a larger extent than a vignette stating that the target's had not control over their body. In Chapter 8 the running on the spot animation may have indicated that the avatar was putting effort into weight-loss, which may have reduced the negative evaluations. Exertion has the advantage of being easy to depict visually; but this might be related to controllability in a mediated relationship. That is, belief in controllability of weight affects negative stereotyping when exertion is perceived. In this model, the fat person is seen to be in control of a controllable element. The study in Chapter 8 would have been a good opportunity to test this mediated model, however The Author did not think of it at the time.

Although body perception and perceived controllability may influence how people respond to avatars, this is not something that is predicted by the Proteus Effect, which is a more general effect. Participants were asked about the extent to which a word was stereotypically associated with a stimulus. This is distinct from a word being associated with a stimulus, which might be worded as 'I think this person is...lazy/determined'.

[20]Although only the first three correlations from column 1 of table 6 would survive the multiple comparison corrections.

A final consideration is the potential role of cognitive dissonance in the evaluations of the animated exemplars and avatars used in this study. Cognitive dissonance is a negative emotional response to the salience of conflicting attitudes, which may lead to a change in attitudes or behaviour to resolve the incongruence (Festinger, 1954). This theory has been used to influence anti-fat attitudes. For example Ciao & Latner (2011) informed participants that their anti-fat attitude did not conform with their self-reported high personal values for kindness and equality. This was associated with a reduction in post-intervention anti-fat attitudes, compared with the control condition. Including measures to test for cognitive dissonance would have been an interesting addition to Chapters 8 and 9. For instance, the shift in attitudes towards larger bodied people may have been greater in those who had stronger existing beliefs about a relationship between body size and fitness, and this may be explained by cognitive dissonance.

Presence, Identification, Relations Arguably, the study in Chapter 9 would have benefitted from measures of self-presence and identification with the avatar, particularly since a bespoke game was produced and had not been rigorously play-tested. These constructs were not measured in any of Yee's original studies, and although subsequent studies have investigated the contribution of presence and identification with avatars towards the Proteus Effect (e.g. Li & Lwin, 2016), few include this as a moderator (see Chapter 4, pages 124, and 134), which is perhaps quite a large oversight (see page 52).

Although identification with ones avatar has been shown to be important for enjoyment of exergames (Li & Lwin, 2016), it is not often used in Proteus Effect research. Identification was discussed in Chapter 5 (from page 169) as a possible mediator between self-similarity and enjoyment but was not included in subsequent studies. An issue with identification is in its unidimensionality. Using an identification measure places on a single continuum between those who identify (share the goals of their avatars etc), or do not. A more promising construct is that of Player-Avatar Relations, since this may predict a wider range of effects

(Chapter 2, page 22).

I have alluded to Player-Avatar Relations throughout this thesis (Banks & Bowman, 2013), but it was not measured in any of the studies featured within. This has potential to be a strong moderator of Proteus-like effects, or the other effects that were addressed in Chapter 4. However, as was discussed in Chapter 2 (page 22, the current measures of PAX-PAR were developed for long term avatar users, such as MMORPG gamers. If the PAX and PAR constructs could be shown to generalise to short term avatar use - e.g. if the propensity to form an Avatar-As-Self attachment, varies across gamers, this could be a valuable construct to investigate. The study in Chapter 9 would have benefitted from such a measure. Including the cPAX in this study would have allowed an exploration of the possible mediating role of Player-Avatar Relations and the Proteus Effect mentioned in Chapter 2 (page 52).

Sample Size and Demographics

Sample size was a considerable limitation throughout the thesis. A number of complicated multivariate method were used, and particularly in Chapter 6 the sample size may have been inadequate for accurately approximating posterior distributions for more factors. This was unavoidable, due to limited resources[21]. Thanks to generous funding from the Manchester Metropolitan University *Stress, Health and Performance* research group, and a research bursary from the British Psychological Society Psychology Postgraduate Affairs Group (PsyPAG)[22] the sample sizes were much larger than they would otherwise have been. Since many of the analyses used Bayesian methods, it would be possible to add more data to increase confidence in the estimates - however, care would be needed in interpreting the factor analysis studies, which would benefit from additional replication and refinement, particularly (as will be discussed below) with more balanced and representative samples.

[21]The Author self-funded one of the studies

[22]For which The Author is incredibly grateful.

As discussed throughout the thesis, the number of participants in each study (with the possible exception of Chapter 5) were smaller than recommended. The consequences of running frequentist statistics on experiments with small samples are well reported (Button et al., 2013). Small samples result in large standard errors, inaccurate estimates of parameters, and as a result higher chances of missing an effect that would reliably occur in the long run (type 2 error), or erroneously observing an effect that would not reoccur in the long run (type 1 error). Such issues extend to factor analytic methods, where one of the aims is to estimate the population covariance parameters amongst groups of items, a process called *parameter recovery*. When estimating these parameters, the True population parameter will either be near the sample parameter (i.e. within the 95% CI), or not. If the True parameter suggests negligible communalities (the degree to which groups of items correlate), and the sample parameter suggests these are strong, we have a type 1 error; whereas the reverse of this would be a type 2 error. When item communalities are high, MacCallum, Widaman, Preacher, & Hong (2001) suggest that parameter recovery will be accurate even in small samples; but when they are low much larger samples (e.g. above the suggested minimum of 400) are required. This may explain the repeated and reliable emergence of the determination factor in multiple datasets, whereas the remaining factors were more susceptible to sample-related variation. In any case, repeated running of the factor-analytic studies with a larger sample would be advisable before forming strong conclusions about the factors. New findings from replications could then be added to the existing, and the Bayesian analyses could be re-run with more confidence in the factor estimates (see Chapter 3, page 90).

In addition to small samples, it has been noted throughout that the demographics of the samples varied considerably over the quantitative chapters; in Chapter 6, the study contributing to the initial construction of the stereotypical structure had substantially more women than men; Chapter 7 had more men than women; and more women than men completed the survey in Chapter 8. This variation may have affected the posterior point estimates, although these remained similar throughout, but may explain the uncertainty in many of the estimates. It may also be problematic that the priors used to inform subsequent studies were constructed from potentially biased estimates. The age ranges also varied between the survey studies, with Chapter 6 having the most limited distribution. Variations in anti-fat attitudes amongst genders have been reported, for instancee O'Brien, Hunter, Halberstadt, & Anderson (2007) found that men scored higher than women on anti-fat subscales measuring *Dislike* of fat people; this could feasibly affect warmth evaluations through responses to affability and arrogance words and would imply that the scales are non-invariant to gender. This warrants further investigation under more balanced conditions with larger sample sizes. This appears to be a more widespread problem in anti-fat research. I checked the first page of Google Scholar search results using the search terms 'Anti-fat attitudes gender' and looked at the gender splits between males and females (all had binary options). All 9 of the studies reported a gender imbalance (although this was only by one participant in one case), with the percentage differences ranging from 1-78%. It is possible that this is a due to the gender imbalances in undergraduate courses (a major source of research participation) discussed in Chapter 6 (page 197), but may also represent a response bias - with women being more likely to participate in body-related studies.

More than three times the number of women than men participated in the study in Chapter 9. In Chapter 2 (page 14) and Chapter 3 (page 71), the target population was defined as people in the UK who are between the ages of 16-44 (representing a large portion of the estimated number of gamers in the UK), and might enjoy exercising using videogames.

This demographic variation is also problematic, given that Chapters 7 to 9 were replication-extension studies. Zwaan et al. (2018) identify a range of purposes that replication may serve, ranging from fraud detection to out-of-sample generalisability. In a close replication, where (for example) the purpose is to detect false positives, the sample and methods ought to be kept as similar to the original study as possible. The studies Chapters 7 and 8 in this thesis are closer to conceptual replication and extension studies, in which different methods are used to observe a similar effect. A weakness of the implementation in these studies was that the demographics between these replications and the original (Chapter 6) were vastly different. This difference was unavoidable given the constraints on the project, and additional work will be required to resolve this weakness. The study in Chapter 9 was framed as a replication and extension of Li et al. (2014), however the participants were quite different from this study, who were ”overweight” children from Singapore. Instead, adults in the UK, who were not selected based on their body-type participated in the Chapter 9 study. Arguably, given the assumption that the Proteus Effect should occur whenever a stereotype is present, that similar effects have been observed in adults (e.g. the Peña studies from Chapter 4), and the large prevalence of weight stigma across ages and many cultures (see Chapter 2, page 44), the effect should occur regardless of the demographics of the participant. Ad hoc suggestions as to why an effect did not replicate would place the Proteus Effect on a degenerating research line (Chapter 3, page 67).

The Research Question

Upon re-reading the original advert for this PhD, the final project bears little resemblance to what was initially proposed (see https://osf.io/hp5a9/). The proposal was for a health-related intervention that used self-similar avatars, with bodies that were either ideal or reflected the participant’s actual body. However, the systematic review in Chapter 4 noted that the research corpus for Self-Discrepancy Theory related approaches to avatar interventions made inconsistent predictions. As such, time was not right for this research line, and another thesis would no doubt be needed to approach these questions. The work on avatar body type, and how it may affect in-game performance was slightly more developed, raised interesting and questionable points, and had a theoretical foundation that could generate predictions. Further, the Proteus Effect has been applied to various behaviours, and in-game exercise was a useful direct measure. However, there were also issues with the methodological rigour of previous studies that were both challenging and enjoyable to address. So, although the absence of an intervention is disappointing, I feel that the state of the avatar-self perception is in a slightly better state. Moving forward, it would be interesting to address the issues with avatar-SDT research - perhaps this could be a future PhD project.

Qualitative Study

The interviews for the Qualitative study in Chapter 5 were conducted by The Author, a white man who, at the time, was in his early 30s and had an athletic build. This may have influenced how the participants discussed their own bodies and their ideals. Indeed, research discussed earlier suggested that pairings with athletic-looking people may result in social comparison and state body dissatisfaction (Wasilenko et al., 2007). As such, participants may have masked their true attitudes and beliefs. Moreover, the purpose of the study may have been obvious, and they may have then responded in what they perceived to be socially desirable. This latter point was partly evident in the participant's responses to the game itself. The Author noted that the participants seemed to hold back their criticism, and on reflection felt that they may have believed that he himself had made the game himself. Comments from a reviewer of an early version of the manuscript version of Chapter 5[23] noted that the questions seemed leading. This may have been the case, and readers are invited to refer to the appendices (https://osf.io/he6us/) to decide for themselves. The Author was inexperienced in Qualitative methods at the time, although received guidance from the supervisory team. There were times where The Author felt he actually was leading participants, but this was after the main items had been covered,

[23] Available at https://psyarxiv.com/6ensp/, and also on the digital appendix page.

and these extended accounts were not included in the analysis.

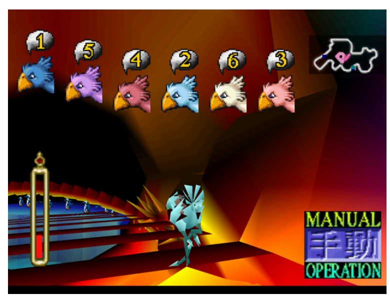
Animation Study

The idea for using animations as a manipulation in Chapter 8 stemmed, in part, from the Chocobo racing sub-game in Final Fantasy VII[24].

(a) A terrible Chocobo.

(b) A great Chocobo.

(c) A wonderful Chocobo.



(d) A Chocobo race.

Figure 10.3: Chocobos from Final Fantasy VII.

In this mini-game, the player must travel around the world collecting Chocobos[25] from various locations and sending them back to an in-game ranch to be evaluated and trained. The attributes of the Chocobos that the player collects are revealed by a non-playable character named Chocobo-Billy who uses descriptors such as “Wonderful”, “Great”, “Good”, or “Terrible”. Based on these descriptors, the player may decide whether to keep the Chocobo to race it or release it into the wild. Even before Chocobo-Billy gives his evaluation, the player can immediately determine whether the Chocobo will eventually be a

[24] A Squaresoft/Square Enix game from 1996, and The Author’s all-time favourite videogame - in fact he will probably be playing the much anticipated Remake as this thesis is being reviewed.

[25] A fictional species of large bird-like bipeds that may be mounted like a horse and raced.

good racer by its behaviour. The “Terrible” Chocobos can be seen plodding with their heads down, already looking defeated, whereas the ‘Wonderful’ Chocobos are presented wings back, eyes forward, and sprinting![26]. These animations powerfully portray the aptitude of the Chocobo, despite identical physical appearances. As such, it was interesting (and scientifically valid, given the research on counter-stereotypical information) to explore how similar animations could affect evaluations of humanoid exemplars.

When choosing the animations for Chapter 8, a variety of animations were tested. These included a *treadmill run*, a normal running, and a sprint. One of the supervisors pointed out that the sprinting exemplar looked a little phrenetic, and so a normal running animation was chosen. For the control condition, a neutral *idle* rather than a negative animation, such as that used to portray “Terrible” Chocobos, was chosen to contrast the *confidently running on the spot* animation.

10.5 Future Research

Since all of the scripts, surveys, and data are available, replicating the studies from this programme *should* be seamless. One direction that is already underway is replicating Chapters 7 and 8 with racially and ethnically diverse exemplars. The body compositions have been kept constant, but skin colour and facial features have been altered to represent Black, Asian, and Hispanic individuals. White avatars were used in the current thesis because of the assumption that there would be negative intersectional stereotypes related to the exemplars, but the replication attempt will explore this systematically. It will be particularly interesting to see if intersectional stereotypes such as “natural black athlete” (Sailes, 1993) are replicated using the current methodology.

There are two tracks that would be interesting continuations of this project. The first (already in progress) is an exploration of stereotypes associated

[26] There are certainly similarities between the presentation of these Chocobos and how the exemplars were presented, for instance the black background.

with race and body-type and would look to confirm whether the intersectional stereotypes discussed in Chapter 2 are still prevalent. A second track would be to look at further improving evaluations of larger bodied people and athletic people. For example, research could be done on the effect of mixed-body type group exercise on person perception by presenting scenes of exemplars with different body types exercising. This may increase the perceived competence of the larger bodied exemplars, and the sociability of the athletic exemplars. The aim here would be to maximise the persuasive features of health advertising and interventional materials. Further, a closer look at gender representations would be a useful approach, particularly given potential differences in the internalisation of athleticism between men and women (this was not studied in this thesis but see Chapter 7 page 253 for a more extensive discussion of this) (Schaefer, Harriger, Heinberg, Soderberg, & Kevin Thompson, 2017; Thompson & Stice, 2001).

A criticism of previous studies into the Proteus Effect and body type was that not enough varieties of body type were explored. This same criticism can also be applied to the current programme, since only larger-bodied and athletic exemplars were used. The Daz Genesis2 model offers a wide variety of manipulations, and so a more incremental approach could be taken. This would potentially require a vast sample size, and/or a within subjects design, but would allow a closer look at how evaluations change with a wider range of forms. In such a study, a range of different body sizes, much like those sent to the body image experts in Chapter 6 could be presented with a shortened set of words, and the evaluations could be compared within participants.

A future direction could be to include information from the other AST systems, such as affective responses, or personality. The former could be manipulated using facial affect; DAZ3D allows for facial expressions to be added to avatars, so they could be depicted as happy, elated, sad, in pain, angry. Facial expression may affect the morality/warmth evaluations. For personality, this could be manipulated using either a scripted introduction, or an attribute summary. An example of the latter is the SPECIAL system from the Fallout series. It stands for Strength, Perception, Endurance, Charisma, Intelligence, Agility, Luck. When creating a character, players assign points to each of these attributes (see figure 10.4).

Figure 10.4: Character creation screen from Fallout

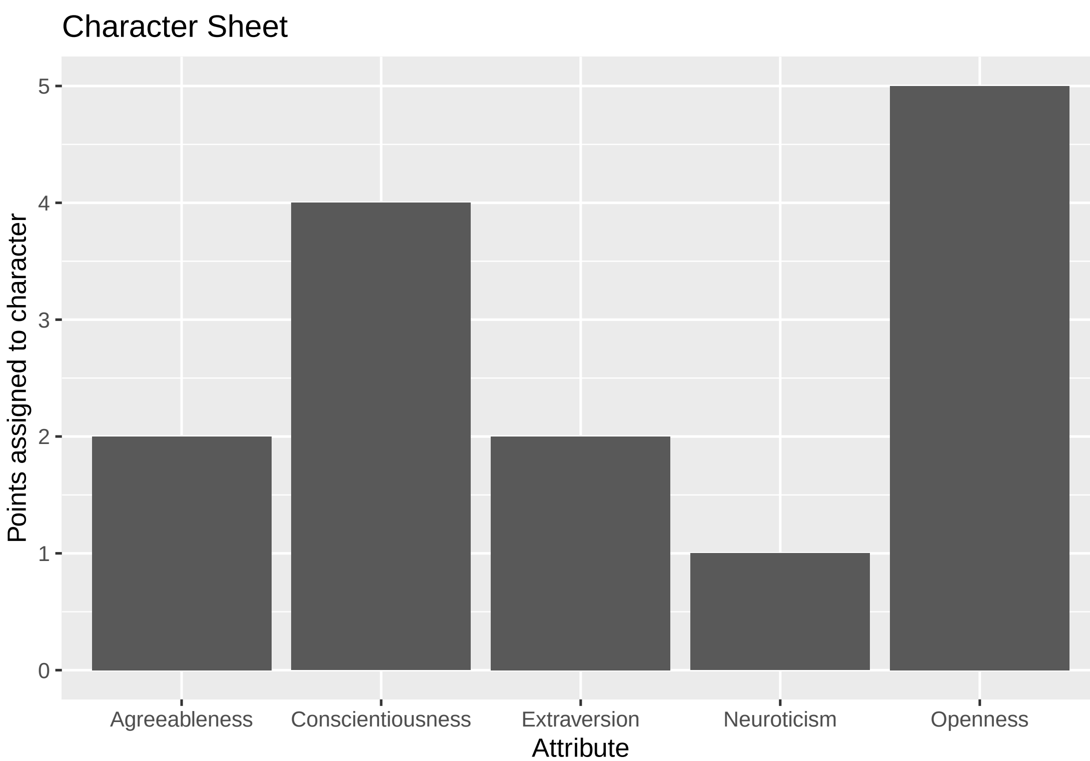


Figure 10.5: A possible method for providing personality-related information to participants. The expected impression formed of such an exemplar would include high conscientiousness and openness.

There are similar functions in many videogames (and role playing games - for instance Dungeons and Dragons). A similar graphic to that shown in figure 10.5 could be presented to participants. The graph is a histogram of the number of points assigned to each of the Big Five personality attributes. If such a manipulation worked, one would expect ratings to be more positive on the competence dimensions when presented with larger bodied avatars.

Following up on section 10.4, it would be interesting to see how negative behavioural animation affects evaluations of athletic-looking individuals, since research suggests that there is an asymmetry in effects, and these should not affect evaluations as much as implied morality would (Paolini & McIntyre, 2019).

10.6 Conclusion

This thesis has demonstrated that using social psychological theory, weight stigmatising evaluations of larger bodied exemplars may be reduced. This knowledge can be used to inform principled character design and character-design systems and applied to both playable and non-playable characters in videogames. There was little to no evidence of the Proteus Effect on in game performance, although an interesting observation was the possible tendency for people to neglect base-rates when using athletic avatars[27].

I found that opinions surrounding athletic and plus-sized bodies were more complicated than the diametric opposites assumed by previous research (Robertson & Vohora, 2008), and that they somewhat contradicted recommendations from Human Computer Interaction researchers to limit body size affordances available to novice avatar designers. These findings led to a structured research programme that aimed to define not just weight stigma, but also investigate an apposite construct for athletic bodied individuals. I found that both groups elicited ambivalent stereotypes, implying that there was more to the *fitness* versus *fatness* story. I was next able to manipulate negative

[27] If this is the case in the current study on exergames, a future direction might be to assess the role of avatars in online gambling. Could a more confident/competent looking avatar lead participants to make riskier decisions during poker matches. This is severely tangential, which is why this has been relegated to a footnote.

evaluations of larger-bodied individuals using richer portrayals, representing the group as active and more competent using counter-stereotypical animations. I finally demonstrated that these enriched representations may affect in game performance; and that the use of athletic avatars may lead players to ignore task-relevant information, which may pose an issue when trying to motivate people to play exergames[28].

10.7 Closing Remarks

In summary, where providing behavioural information does not cancel out weight stigma towards exemplars, it seems to reduce it. Rather than encouraging people to select "normal" or athletic avatars, focus should be shifted to creating more diffuse and positive representations of a wider variety of body types. This would allow players to explore different bodies, whilst not feeling ashamed when exploring virtual worlds with their own.

[28]For instance, an athletic avatar may inspire someone to skip straight to a difficult setting, leading to them being repeatedly beaten, resulting in a reduction of self-efficacy, and less enjoyment.

References

2019 essential facts about the computer and video game industry - entertainment software association. (n.d.).

Adam, H., & Galinsky, A. D. (2012). Enclothed cognition. *Journal of Experimental Social Psychology*, *48*(4), 918–925.

Adler, P., & Adler, P. A. (1987). Role conflict and identity salience: College athletics and the academic role. *The Social Science Journal*, *24*(4), 443–455.

Adobe. (2018). Mixamo. Retrieved from https://www.mixamo.com/

Ahn, S. J. (2015a). Incorporating immersive virtual environments in health promotion campaigns: A construal level theory approach. *Health Communication*, *30*(6), 545–556.

Ahn, S. J. (2015b). Using avatars and agents to promote real-world health behavior changes. *The Digital Patient: Advancing Healthcare, Research, and Education*, 171–180.

Ahn, S. J. (2016). Virtual exemplars in health promotion campaigns. *Journal of Media Psychology*.

Ahn, S. J., & Fox, J. (2017a). *Immersive virtual environments, avatars, and agents for health.*

Ahn, S. J., & Fox, J. (2017b). *Virtual environments, avatars, and agents in health and risk messaging. Oxford Research Encyclopedia of Communication. Ed.* Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.

013.325

Ahn, S. J., Fox, J., & Hahm, J. M. (2014). Using virtual doppelgängers to increase personal relevance of health risk communication. In *International conference on intelligent virtual agents* (pp. 1–12). Springer.

Ahn, S. J., Johnsen, K., Robertson, T., Moore, J., Brown, S., Marable, A., & Basu, A. (2015). Using virtual pets to promote physical activity in children: An application of the youth physical activity promotion model. *Journal of Health Communication*, *20*(7), 807–815.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Ajzen, I. (2002). Constructing a tpb questionnaire: Conceptual and methodological considerations.

Alahäivälä, T., & Oinas-Kukkonen, H. (2016). Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. *International Journal of Medical Informatics*, *96*, 62–70. https://doi.org/http://dx.doi.org/10.1016/j.ijmedinf.2016.02.006

Alciati, A. (2004). *A book of emblems: The emblematum liber in latin and english.* McFarland.

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Chang, W. (2018). *Rmarkdown: Dynamic documents for r.* Retrieved from https://CRAN.R-project.org/package=rmarkdown

Alleva, J. M., Diedrichs, P. C., Halliwell, E., Martijn, C., Stuijfzand, B. G., Treneman-Evans, G., & Rumsey, N. (2018). A randomised-controlled trial investigating potential underlying mechanisms of a functionality-based approach to improving women's body image. *Body Image*, *25*, 85–96.

Alleva, J. M., Martijn, C., Van Breukelen, G. J., Jansen, A., & Karos, K. (2015). Expand your horizon: A programme that improves body image and reduces self-objectification by training women to focus on body functionality. *Body Image*, *15*, 81–89.

Andrews, M., & Baguley, T. (2013). Prior approval: The growth of bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 1–7.

Anesbury, T., & Tiggemann, M. (2000). An attempt to reduce negative stereotyping of obesity in children by changing controllability beliefs. *Health Education Research*, *15*(2), 145–152.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, *40*(4), 471–499.

Arnett, J. J. (2016). The neglected 95%: Why american psychology needs to become less american.

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Bacon, J. G., Scheltema, K. E., & Robinson, B. E. (2001). Fat phobia scale revisited: The short form. *International Journal of Obesity*, *25*(2), 252.

Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics*, *35*(2), 73–80.

Bailenson, J. N., Iyengar, S., Yee, N., & Collins, N. A. (2008). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly*, *72*(5), 935–961.

Bailey, B., & McInnis, K. (2011). Energy cost of exergaming: A comparison of the energy cost of 6 forms of exergaming. *Archives of Pediatrics & Adolescent Medicine*, *165*(7), 597–602. https://doi.org/10.1001/archpediatrics.2011.15

Ball, K., Crawford, D., & Owen, N. (2000). Obesity as a barrier to physical activity. *Australian and New Zealand Journal of Public Health*, *24*(3), 331–333.

Banakou, D., Kishore, S., & Slater, M. (2018). Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in Psychology*, *9*, 917.

Banks, J. (2007). *Human-technology relationality and self-network organization: Players and avatars in world of warcraft* (PhD thesis). Colorado State University. Libraries.

Banks, J., & Bowman, N. D. (2013). Close intimate playthings? Understanding player-avatar relationships as a function of attachment, agency, and intimacy. *Selected Papers of Internet Research*, *3*.

Banks, J., & Bowman, N. D. (2016a). Avatars are (sometimes) people too: Linguistic indicators of parasocial and social ties in player–avatar relationships. *New Media & Society*, *18*(7), 1257–1276.

Banks, J., & Bowman, N. D. (2016b). Emotion, anthropomorphism, realism, control: Validation of a merged metric for player–avatar interaction (pax). *Computers in Human Behavior*, *54*, 215–223.

Banks, J., Bowman, N. D., Lin, J.-H. T., Pietschmann, D., & Wasserman, J. A. (2019). The common player-avatar interaction scale (cPAX): Expansion and cross-language validation. *International Journal of Human-Computer Studies*, *129*, 64–73.

Bargh, J. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. Inr. S. Wyer, jr., & tk srull (eds.), handbook ofsocial cognition (pp. 1-40). Hillsdale, NJ: Erlbaum.

Barlösius, E., & Philipps, A. (2015). Felt stigma and obesity: Introducing the generalized other. *Social Science & Medicine*, *130*, 9–15.

Barry, V. W., Baruth, M., Beets, M. W., Durstine, J. L., Liu, J., & Blair, S. N. (2014). Fitness vs. Fatness on all-cause mortality: A meta-analysis. *Progress in Cardiovascular Diseases*, *56*(4), 382–390.

Behm-Morawitz, E., Lewallen, J., & Choi, G. (2016). A second chance at health: How a 3D virtual world can improve health self-efficacy for weight loss management among adults, *19*(2), 74–79. https://doi.org/10.1089/cyber.2015.0317

Belk, R. (2011). Benign envy. *AMS Review*, *1*(3-4), 117–134.

Bem, D. J. (1966). Inducing belief in false confessions. *Journal of Personality and Social Psychology*, *3*(6), 707.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, *6*, 1–62.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407.

Bem, D. J., & McConnell, H. K. (1970). Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, *14*(1), 23.

Benefits of exercise. (2018, June). Retrieved from https://www.nhs.uk/live-well/exercise/exercise-health-benefits/

Berlin, J. A., & Colditz, G. A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *American Journal of Epidemiology*, *132*(4), 612–628.

Bhaskar, R. (2013). *A realist theory of science*. Routledge.

Bhaskar, R. (2016). *Enlightened common sense: The philosophy of critical realism*. Routledge.

Bidart, F. (2003). Borges and i. *Great American Prose Poems: From Poe to the Present*, 132–134.

Biddiss, E., & Irwin, J. (2010). Active video games to promote physical activity in children and youth: A systematic review. *Archives of Pediatrics & Adolescent*

Medicine, *164*(7), 664–672.

Birk, M., Atkins, C., Bowey, J., & Mandryk, R. (2016). Fostering intrinsic motivation through avatar identification in digital games. In (pp. 2982–2995). ACM.

Black, M. J., Sokol, N., & Vartanian, L. R. (2014). The effect of effort and weight controllability on perceptions of obese individuals. *The Journal of Social Psychology*, *154*(6), 515–526.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828.

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality: A prospective study of healthy and unhealthy men. *Jama*, *273*(14), 1093–1098.

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Jama*, *262*(17), 2395–2401.

Blastland, M. (2019). *The hidden half: How the world conceals its secrets*. Atlantic Books.

Bock, B. C., Dunsiger, S. I., Ciccolo, J. T., Serber, E. R., Wu, W.-C., Tilkemeier, P., … Marcus, B. H. (2019). Exercise videogames, physical activity, and health: Wii heart fitness: A randomized clinical trial. *American Journal of Preventive Medicine*, *56*(4), 501–511.

Bogost, I. (2007). *Persuasive games: The expressive power of videogames*. Mit Press.

Bolton, G. E. (1991). A comparative model of bargaining: Theory and evidence. *The American Economic Review*, 1096–1136.

Bonett, D. G. (2012). Replication-extension studies. *Current Directions in Psychological Science*, *21*(6), 409–412. https://doi.org/10.1177/

0963721412459512

Borges, J. L., & Hurley, A. (1998). *Fictions*. Penguin Classics.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.

Bradley, P. (2018). Bots and data quality on crowdsourcing platforms. Retrieved from https://blog.prolific.co/bots-and-data-quality-on-crowdsourcing-platforms/

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.

Brewis, A. A. (2014). Stigma and the perpetuation of obesity. *Social Science & Medicine*, *118*, 152–158.

Brewis, A., SturtzSreetharan, C., & Wutich, A. (2018). Obesity stigma as a globalizing health challenge. *Globalization and Health*, *14*(1), 20.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, *76*(1), 15.

Brown, S. A. (2005). Measuring perceived benefits and perceived barriers for physical activity. *American Journal of Health Behavior*, *29*(2), 107–116.

Buerkner, P.-C. (2016). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*.

Buisine, S., Guegan, J., Barré, J., Segonds, F., & Aoussat, A. (2016). Using avatars to tailor ideation process to innovation strategy. *Cognition, Technology & Work*, *18*(3), 583–594.

Burns, D. M., Fox, E. B., Greenstein, M., & Montgomery, D. A. (2019). An old task in new clothes: A preregistered direct replication attempt of enclothed cognition effects on stroop performance.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (n.d.). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*.

Bürkner, P.-C., & Vuorre, M. (2018). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2515245918823199.

Cacioli, J.-P., & Mussap, A. J. (2014). Avatar body dimensions and men's body image. *Body Image*, *11*(2), 146–155. https://doi.org/10.1016/j.bodyim.2013.11.005

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, *35*(1), 38.

Carlston, D. E. (1994). *Associated systems theory: A systematic approach to cognitive representations of persons.* Lawrence Erlbaum Associates, Inc.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126.

Castleberry, A. (2014). NVivo 10 [software program]. Version 10. QSR international; 2012. AJPE.

Castronova, E. (2003). Theory of the avatar.

Chalabaev, A., Sarrazin, P., Fontayne, P., Boiché, J., & Clément-Guillotin, C. (2013). The influence of sex stereotypes and gender roles on participation and performance in sport and exercise: Review and future directions. *Psychology of Sport and Exercise*, *14*(2), 136–144.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.

Chirkov, V., & Anderson, J. (2018). Statistical positivism versus critical scientific realism. A comparison of two paradigms for motivation research: Part 1. A philosophical and empirical analysis of statistical positivism. *Theory & Psychology*, *28*(6), 712–736.

Ciao, A. C., & Latner, J. D. (2011). Reducing obesity stigma: The effectiveness of cognitive dissonance and social consensus interventions. *Obesity*, *19*(9), 1768–1774.

Cole, J. M., & Grogan, S. (2019). "Kind of like a barbie doll, but for grown men!": Women gamers' accounts of female bodies in digital games. *Psychology of Women and Equalities Review*.

Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, *119*(1), 51.

Coram, S. (2007). Race formations (evolutionary hegemony) and theAping'of the australian indigenous athlete. *International Review for the Sociology of Sport*, *42*(4), 391–409.

Craft, L. L., & Landers, D. M. (1998). The effect of exercise on clinical depression and depression resulting from mental illness: A meta-analysis. *Journal of Sport and Exercise Psychology*, *20*(4), 339–357.

Crawford, J. T., Jussim, L., Madon, S., Cain, T. R., & Stevens, S. T. (2011). The use of stereotypes and individuating information in political person perception. *Personality and Social Psychology Bulletin*, *37*(4), 529–542.

Crawford, S., & Eklund, R. C. (1994). Social physique anxiety, reasons for exercise, and attitudes toward exercise settings. *Journal of Sport and Exercise Psychology*, *16*(1), 70–82.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

Crusius, J., & Lange, J. (2014). What catches the envious eye? Attentional biases within malicious and benign envy. *Journal of Experimental Social Psychology*, *55*, 1–11.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The bias map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631.

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J.-P., … others. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33.

Cui, J., Aghajan, Y., Lacroix, J., Halteren, A. van, & Aghajan, H. (2009). Exercising at home: Real-time interaction and experience sharing using avatars. *Entertainment Computing*, *1*(2), 63–73.

Cunningham, S. (2018). Causal inference: The mixtape. Manuscript.

Czopp, A. M. (2010). Studying is lame when he got game: Racial stereotypes and the discouragement of black student-athletes from schoolwork. *Social Psychology of Education*, *13*(4), 485–498.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt.

D'Anastasio, C. (2020). Coronavirus fears spark a run on nintendo's ring fit adventure. Retrieved from https://www.wired.com/story/ring-fit-shortage-coronavirus-covid-19/

Darville, G., Anderson –Lewis, C., Stellefson, M., Lee, Y.-H., MacInnes, J., Pigg, R. M., … Thomas, S. (2018). Customization of avatars in a hpv digital gaming intervention for college-age males: An experimental study. *Simulation & Gaming*, *49*(5), 515–537. https://doi.org/10.1177/1046878118799472

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*(5), 800.

De Caroli, M. E., & Sagone, E. (2013). Anti-fat prejudice and stereotypes in psychology university students. *Procedia-Social and Behavioral Sciences*, *84*, 1184–1189.

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. Retrieved from http://www.jstatsoft.org/v64/i04/

Diedrichs, P. C., & Lee, C. (2010). GI joe or average joe? The impact of average-size and muscular male fashion models on men's and women's body image and advertisement effectiveness. *Body Image*, *7*(3), 218–226.

Diedrichs, P. C., & Lee, C. (2011). Waif goodbye! Average-size female models promote positive body image and appeal to consumers. *Psychology & Health*, *26*(10), 1273–1291.

Diekfuss, J. A. (2012). *The dumb jock: How stereotype threats impact student-athletes' standards of academic competence* (PhD thesis).

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*(10), 1171–1188.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference.* Macmillan International Higher Education.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*(1), 417–440.

Drasgow, F. (2004). Polychoric and polyserial correlations. *Encyclopedia of Statistical Sciences*, *9*.

Ducheneaut, N., Wen, M.-H., Yee, N., & Wadley, G. (2009). Body and mind: A study of avatar personalization in three virtual worlds. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1151–1160). ACM.

Dunaev, J. L., Brochu, P. M., & Markey, C. H. (2018). Imagine that! The effect of counterstereotypic imagined intergroup contact on weight bias. *Health Psychology*, *37*(1), 81.

Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, *30*(4), 510.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., … others. (2016). Many labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82.

Elmagarmid, A., Fedorowicz, Z., Hammady, H., Ilyas, I., Khabsa, M., & Ouzzani, M. (2014). Rayyan: A systematic reviews web app for exploring and filtering searches for eligible studies for cochrane reviews. In *Evidence-informed public health: Opportunities and challenges* (p. 9). Hyderabad, India: John Wiley & Sons.

Epskamp, S., & Nuijten, M. (2015). Statcheck: Extract statistics from articles and recompute p values (r package version 1.0. 1.).

Estimating the reproducibility of psychological science. (2015). *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Etz, A., & Vandekerckhove, J. (2018). Introduction to bayesian inference for psychology. *Psychonomic Bulletin & Review*, *25*(1), 5–34.

Falbén, J. K., Tsamadi, D., Golubickis, M., Olivier, J. L., Persson, L. M., Cunningham, W. A., & Macrae, C. N. (2019). Predictably confirmatory: The influence of stereotypes during decisional processing. *Quarterly Journal of Experimental Psychology*, 1747021819844219.

Feinerer, I., & Hornik, K. (2018). *Tm: Text mining package*. Retrieved from https://CRAN.R-project.org/package=tm

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in r. *Journal of Statistical Software*, *25*(5), 1–54. Retrieved from http://www.jstatsoft.org/v25/i05/

Ferguson, C. J. (2007). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior*, *12*(4), 470–482.

Ferguson, C. J. (2013). In the eye of the beholder: Thin-ideal media affects some, but not most, viewers in a meta-analytic review of body dissatisfaction in women and men. *Psychology of Popular Media Culture*, *2*(1), 20.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Field, S. M., Hoekstra, R., Bringmann, L., & Ravenzwaaij, D. van. (2019). When and why to replicate: As easy as 1, 2, 3? *Collabra: Psychology*, *5*(1).

Finco, M. D., Reategui, E., Zaro, M. A., Sheehan, D. D., & Katz, L. (2015). Exergaming as an alternative for students unmotivated to participate in regular physical education classes. *International Journal of Game-Based Learning (IJGBL)*, *5*(3), 1–10.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research.*

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878.

Fiske, S. T., & Taylor, S. E. (1991). Social cognition, 2nd. *NY: McGraw-Hill*, 16–15.

Flake, J. K., & Fried, E. I. (2019). Measurement schmeasurement: Questionable measurement practices and how to avoid them.

Flint, S. W., Hudson, J., & Lavallee, D. (2013). Counter-conditioning as an intervention to modify anti-fat attitudes. *Health Psychology Research*, *1*(2).

Flint, S. W., Hudson, J., & Lavallee, D. (2016). The portrayal of obesity in uk national newspapers. *Stigma and Health*, *1*(1), 16.

Flint, S. W., & Reale, S. (2018). Weight stigma in frequent exercisers: Overt, demeaning and condescending. *Journal of Health Psychology*, *23*(5), 710–719.

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *49*(2), 78.

Ford, R., & Cole, J. (2019). "I don't want to be a fairy princess": Qualitative accounts of female gamers' avatar choices. *Psychology of Women and Equalities Review*.

Fox, J. (2010). The use of virtual self models to promote self-efficacy and physical activity performance.

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction*, *30*(5), 401–432. https://doi.org/10.1080/07370024.2014.921494

Fox, J., Bailenson, J., & Binney, J. (2009a). Virtual experiences, physical behaviors: The effect of presence on imitation of an eating avatar. *Presence: Teleoperators and Virtual Environments*, *18*(4), 294–303.

Fox, J., Bailenson, J., & Binney, J. (2009b). Virtual experiences, physical behaviors: The effect of presence on imitation of an eating avatar. *Presence: Teleoperators and Virtual Environments*, *18*(4), 294–303.

Fox, J., & Bailenson, J. N. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, *12*(1), 1–25.

Fox, J., Bailenson, J. N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, *29*(3), 930–938. https://doi.org/http://dx.doi.org/10.1016/j.chb.2012.12.027

Frank, M. G., & Gilovich, T. (1988). The dark side of self-and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, *54*(1), 74.

Frederick, C. M., & Morrison, C. S. (1996). Social physique anxiety: Personality constructs, motivations, exercise attitudes, and behaviors. *Perceptual and Motor Skills*, *82*(3), 963–972. https://doi.org/10.2466/pms.1996.82.3.963

Freijy, T., & Kothe, E. J. (2013). Dissonance-based interventions for health behaviour change: A systematic review. *British Journal of Health Psychology*, *18*(2), 310–337.

Galvan Debarba, H., Molla, E., Herbelin, B., & Boulic, R. (2015). Characterizing embodied interaction in first and third person perspective viewpoints. In *2015 ieee symposium on 3D user interfaces (3DUI)*.

Gapinski, K. D., Schwartz, M. B., & Brownell, K. D. (2006). Can television change anti-fat attitudes and behavior? 1. *Journal of Applied Biobehavioral Research*, *11*(1), 1–28.

Garn, A. C., Baker, B. L., Beasley, E. K., & Solmon, M. A. (2012). What are the benefits of a commercial exergaming platform for college students? Examining physical activity, enjoyment, and future intentions. *Journal of Physical Activity and Health*, *9*(2), 311–318.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*.

Gibson, C. E., Losee, J., & Vitiello, C. (2014). A replication attempt of stereotype susceptibility (shih, pittinsky, & ambady, 1999): Identity salience and shifts in quantitative performance. *Social Psychology*, *45*(3), 194.

Gill, R. (2012). Media, empowerment and the "sexualization of culture" debates. *Sex Roles*, *66*(11), 736–745. https://doi.org/10.1007/s11199-011-0107-1

Ginis, K. M., & Bassett, R. (2012). Exercise: Effects on body image. In *Encyclopedia of body image and human appearance* (pp. 412–417). Elsevier.

Goretzko, D., Pham, T. T. H., & Bühner, M. (2019). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 1–12.

Greenleaf, C., Chambliss, H., Rhea, D. J., Martin, S. B., & Morrow Jr, J. R. (2006). Weight stereotypes and behavioral intentions toward thin and fat peers among white and hispanic adolescents. *Journal of Adolescent Health*, *39*(4), 546–552.

Greenleaf, C., Klos, L., Hauff, C., Hennum, A., Pozoliniski, B., & Serafin, G. (2019). "Unless you puke, faint, or die, keep going!" Exploring weight stigma in the gym on the biggest loser. *Fat Studies*, *8*(2), 110–126.

Greenleaf, C., & Weiller, K. (2005). Perceptions of youth obesity among physical educators. *Social Psychology of Education*, *8*(4), 407–423.

Grogan, S. (2017). *Body image: Understanding body dissatisfaction in men, women and children*. Taylor & Francis.

Grogan, S., Flett, K., Clark-Carter, D., Conner, M., Davey, R., Richardson, D., & Rajaratnam, G. (2011). A randomized controlled trial of an appearance-related smoking intervention. *Health Psychology*, *30*(6), 805.

Gronau, Q. F., & Singmann, H. (2017). *Bridgesampling: Bridge sampling for marginal likelihoods and bayes factors*. Retrieved from https://CRAN.R-project.org/package=bridgesampling

Grubbs, L., & Carter, J. (2002). The relationship of perceived benefits and barriers to reported exercise behaviors in college undergraduates. *Family & Community Health*, *25*(2), 76–84.

Guy, S., Ratzki-Leewing, A., & Gwadry-Sridhar, F. (2011). Moving beyond the stigma: Systematic review of video games and their potential to combat obesity. *International Journal of Hypertension*, *2011*.

Hall, E. E., Baird, S. A., Gilbert, D. N., Miller, P. C., & Bixby, W. R. (2011). Viewing television shows containing ideal and neutral body images while exercising: Does type of body image content influence exercise performance and body image in women? *Journal of Health Psychology*, *16*(6), 938–946. https://doi.org/10.1177/1359105310396394

Halliwell, E. (2013). The impact of thin idealized media images on body satisfaction: Does body appreciation protect women from negative effects? *Body Image*, *10*(4), 509–514. https://doi.org/https://doi.org/10.1016/j.bodyim.2013.07.004

Halliwell, E., Dittmar, H., & Orsborn, A. (2007). The effects of exposure to muscular male models among men: Exploring the moderating role of gym use and exercise motivation. *Body Image*, *4*(3), 278–287. https://doi.org/https://doi.org/10.1016/j.bodyim.2007.04.006

Hardin, E. E., & Lakin, J. L. (2009). The integrated self-discrepancy index: A reliable and valid measure of self-discrepancies. *Journal of Personality Assessment*, *91*(3), 245–253.

Haslam, S. A., Turner, J. C., Oakes, P. J., Reynolds, K. J., & Doosje, B. (2002). From personal pictures in the head to collective tools in the world: How shared stereotypes allow groups to represent and change social reality.

Heine, S. J., & Hamamura, T. (2007). In search of east asian self-enhancement. *Personality and Social Psychology Review*, *11*(1), 4–27.

Hendrick, C. (1990). Replications, strict replications, and conceptual replications: Are they important. *Journal of Social Behavior and Personality*, *5*(4), 41–49.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83.

Heuer, C. A., McClure, K. J., & Puhl, R. M. (2011). Obesity stigma in online news: A visual content analysis. *Journal of Health Communication*, *16*(9), 976–987.

Higgin, T. (2012). *Gamic race: Logics of difference in videogame culture* (PhD thesis). UC Riverside.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319.

Higgins, E. T. (1996). The" self digest": Self-knowledge serving self-regulatory functions. *Journal of Personality and Social Psychology*, *71*(6), 1062.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., … Sterne, J. A. C. (2011). The cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*. https://doi.org/10.1136/bmj.d5928

Hilton, J. L., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. *Journal of Personality and Social Psychology*, *57*(2), 201.

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–Garcia, I., & Vila–Abad, E. (2010). Polychoric versus pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153.

Homan, K., McHugh, E., Wells, D., Watson, C., & King, C. (2012). The effect of viewing ultra-fit images on college women's body dissatisfaction. *Body Image*, *9*(1), 50–56. https://doi.org/https://doi.org/10.1016/j.bodyim.2011.07.006

Horne, M., Hill, A., Murells, T., Ugail, H., Chinnadorai, R., Hardy, M., & others. (2019). Using avatars in weight management settings: A systematic review. *Internet Interventions*, 100295.

Horton, D., & Richard Wohl, R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, *19*(3), 215–229.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Huang, H.-C., Wong, M.-K., Lu, J., Huang, W.-F., & Teng, C.-I. (2017). Can using exergames improve physical fitness? A 12-week randomized controlled trial. *Computers in Human Behavior*, *70*, 310–316.

IAB. (2011). "33 million uk game players". Retrieved from https://www.iabuk.net/news/33-million-uk-game-players

Ioannidis, J. P. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124. Retrieved from https://doi.org/10.1371/journal.pmed.0020124

Jette, M., Sidney, K., & Blümchen, G. (1990). Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical Cardiology*, *13*(8), 555–565.

Jin, S.-A. A. (2009). Avatars mirroring the actual self versus projecting the ideal self: The effects of self-priming on interactivity and immersion in an exergame, wii fit. *CyberPsychology & Behavior*, *12*(6), 761–765.

Jin, S.-A. A. (2012). The virtual malleable self and the virtual identity discrepancy model: Investigative frameworks for virtual possible selves and others in avatar-based identity construction and social interaction. *Computers in Human Behavior*, *28*(6), 2160–2168.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Joo, Y. K., & Kim, K. (2017). When you exercise your avatar in a virtual game: The role of avatars' body shape and behavior in users' health behavior. *Interacting with Computers*, *29*(3), 455–466.

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in) accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, *24*(6), 490–497.

Kafai, Y. B., Cook, M. S., & Fields, D. A. (2007a). " Blacks [sic] deserve bodies too!" Design and discussion about diversity and race in a tween online world. In *DiGRA conference*.

Kafai, Y. B., Fields, D. A., & Cook, M. S. (2007b). Your second selves: Resources, agency, and constraints in avatar designs and identity play in a tween virtual world. In *DiGRA conference*.

Kafai, Y. B., Fields, D. A., & Cook, M. S. (2010). Your second selves: Player-designed avatars. *Games and Culture*, *5*(1), 23–42.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237.

Kail, P. J. E. (2019). Epistemology: The key thinkers. In S. Hetherington (Ed.). Bloomsbury Publishing.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark iv. *Educational and Psychological Measurement*, *34*(1), 111–117.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*(3), 280.

Kaye, L. K., Pennington, C. R., & McCann, J. J. (2018). Do casual gaming environments evoke stereotype threat? Examining the effects of explicit priming and avatar gender. *Computers in Human Behavior*, *78*, 142–150.

Kelly, K. (1998). The third culture. *Science*, *279*(5353), 992–993.

Khabsa, M., Elmagarmid, A., Ilyas, I., Hammady, H., & Ouzzani, M. (2015). Learning to identify relevant studies for systematic reviews using random forest and external information. *Machine Learning*, 1–18. https://doi.org/10.1007/s10994-015-5535-7

Kim, Y., & Sundar, S. S. (2012). Visualizing ideal self vs. Actual self through avatars: Impact on preventive health outcomes. *Computers in Human Behavior*, *28*(4), 1356–1364. https://doi.org/http://dx.doi.org/10.1016/j.chb.2012.02.021

King, C. (2004). Race and cultural identity: Playing the race game inside football. *Leisure Studies*, *23*(1), 19–30.

King, N., Horrocks, C., & Brooks, J. (2018). *Interviews in qualitative research*. SAGE Publications Limited.

King, S. (2006). *The dark tower vii: The dark tower* (Vol. 7). Hachette UK.

Klimmt, C., Hartmann, T., & Schramm, H. (2006). Parasocial interactions and relationships. *Psychology of Entertainment*, 291–313.

Kolko, B. E. (1999). Representing bodies in virtual space: The rhetoric of avatar design. *The Information Society*, *15*(3), 177–186.

Kowert, R., Festl, R., & Quandt, T. (2014). Unpopular, overweight, and socially inept: Reconsidering the stereotype of online gamers. *Cyberpsychology, Behavior, and Social Networking*, *17*(3), 141–146.

Kowert, R., Griffiths, M. D., & Oldmeadow, J. A. (2012). Geek or chic? Emerging stereotypes of online gamers. *Bulletin of Science, Technology & Society*, *32*(6), 471–479.

Kruschke, J. K., & Liddell, T. M. (2017). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*, 1–29.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284.

Kuo, H.-C., Lee, C.-C., & Chiou, W.-B. (2016). The power of the virtual ideal self in weight control: Weight-reduced avatars can enhance the tendency to delay gratification and regulate dietary practices. *CyberPsychology, Behavior & Social Networking*, *19*(2), 80–85. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=113040553&site=ehost-live

Lakatos, I., Worrall, J., & Currie, G. (1979). The methodology of scientific research programmes: Philosophical papers.

Lakens, D. (2014). Grounding social embodiment. *Social Cognition*, *32*(Supplement), 168.

Lakens, D. (2015). On the challenges of drawing conclusions from p-values just below 0.05. *PeerJ*, *3*, e1142.

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., … others. (2018). Justify your alpha. *Nature Human Behaviour*, *2*(3), 168.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269.

Lam, J. W., Sit, C. H., & McManus, A. M. (2011). Play pattern of seated video game and active "exergame" alternatives. *Journal of Exercise Science & Fitness*, *9*(1), 24–30.

Lamboglia, C. M. G. F., Silva, V. T. B. L. da, Vasconcelos Filho, J. E. de, Pinheiro, M. H. N. P., Munguba, M. C. da S., Silva Júnior, F. V. I., … Silva, C. A. B. da. (2013). Exergaming as a strategic tool in the fight against childhood obesity: A systematic review. *Journal of Obesity*, *2013*.

Lange, J., & Crusius, J. (2015). The tango of two deadly sins: The social-functional relation of envy and pride. *Journal of Personality and Social Psychology*, *109*(3), 453.

Lange, J., Weidman, A. C., & Crusius, J. (2018). The painful duality of envy: Evidence for an integrative theory and a meta-analysis on the relation of envy and schadenfreude. *Journal of Personality and Social Psychology*, *114*(4), 572.

LeBlanc, A. G., Chaput, J.-P., McFarlane, A., Colley, R. C., Thivel, D., Biddle, S. J., … Tremblay, M. S. (2013). Active video games and health indicators in children and youth: A systematic review. *PloS One*, *8*(6), e65351.

Lee, C. D., Blair, S. N., & Jackson, A. S. (1999). Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men–. *The American Journal of Clinical Nutrition*, *69*(3), 373–380.

Lee, C. D., Folsom, A. R., & Blair, S. N. (2003). Physical activity and stroke risk: A meta-analysis. *Stroke*, *34*(10), 2475–2481.

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50.

Li, B. J., & Lwin, M. O. (2016). Player see, player do testing an exergame motivation model based on the influence of the self avatar. *Computers in Human Behavior*, *59*, 350–357. https://doi.org/http://dx.doi.org/10.1016/j.

chb.2016.02.034

Li, B. J., Lwin, M. O., & Jung, Y. (2014). Wii, myself, and size: The influence of proteus effect and stereotype threat on overweight children's exercise motivation and behavior in exergames. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, *3*(1), 40–48.

Li, D. D., Liau, A. K., & Khoo, A. (2013). Player–avatar identification in video gaming: Concept and measurement. *Computers in Human Behavior*, *29*(1), 257–263. https://doi.org/http://dx.doi.org/10.1016/j.chb.2012.09.002

Liao, C. (2011). Virtual fashion play as embodied identity re/assembling: Second life fashion bloggers and their avatar bodies. In *Reinventing ourselves: Contemporary concepts of identity in virtual worlds* (pp. 101–127). Springer.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology.*

Limperos, A. M., & Schmierbach, M. (2016). Understanding the relationship between exergame play experiences, enjoyment, and intentions for continued play. *Games for Health Journal*, *5*(2), 100–107.

López-Pérez, B., Ambrona, T., Wilson, E. L., & Khalil, M. (2016). The effect of enclothed cognition on empathic responses and helping behavior. *Social Psychology.*

Luo, Z., Westerman, D., & Banks, J. (2019). Extending the self: Player-avatar relations and presence among us and chinese gamers. *Journal for Virtual Worlds Research*, *12*(3).

Lüdecke, D. (2018). *Sjstats: Statistical functions for regression models (version 0.16.0).* https://doi.org/10.5281/zenodo.1284472

Lwin, M. O., & Malik, S. (2012). The efficacy of exergames-incorporated physical education lessons in influencing drivers of physical activity: A comparison of children and pre-adolescents. *Psychology of Sport and Exercise*, *13*(6), 756–760. https://doi.org/http://dx.doi.org/10.1016/j.psychsport.2012.04.013

Lyons, E. J., & Hatkevich, C. (2013). Prevalence of behavior changing strategies in fitness video games: Theory-based content analysis. *Journal of Medical Internet Research*.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611–637.

Mainous III, A. G., Tanner, R. J., Rahmanian, K. P., Jo, A., & Carek, P. J. (2019). Effect of sedentary lifestyle on cardiovascular disease risk among healthy adults with body mass indexes 18.5 to 29.9 kg/m2. *The American Journal of Cardiology*, *123*(5), 764–768.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542.

Makuch, E. (n.d.). GTA 5 reaches impressive new sales milestone, as microtransaction spending grows. Retrieved from https://www.gamespot.com/articles/gta-5-reaches-impressive-new-sales-milestone-as-mi/1100-6473475/

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760.

Mancini, T., Imperato, C., & Sibilla, F. (2019). Does avatar's character and emotional bond expose to gaming addiction? Two studies on virtual self-discrepancy, avatar identification and gaming addiction in massively multiplayer online role-playing game players. *Computers in Human Behavior*, *92*, 297–305. https://doi.org/https://doi.org/10.1016/j.chb.2018.11.007

Mancini, T., & Sibilla, F. (2017). Offline personality and avatar customisation. Discrepancy profiles and avatar identification in a sample of mmorpg players. *Computers in Human Behavior*, *69*, 275–283. https://doi.org/https://doi.org/10.1016/j.chb.2016.12.031

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954.

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E.-J. (2017). A bayesian bird's eye view of ëReplications of important results in social psychologyí. *Royal Society Open Science*, *4*(1), 160426.

Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology*, *52*(3), 489.

Martins, N., Williams, D. C., Harrison, K., & Ratan, R. A. (2009). A content analysis of female body imagery in video games. *Sex Roles*, *61*(11), 824. https://doi.org/10.1007/s11199-009-9682-9

Martins, N., Williams, D. C., Ratan, R. A., & Harrison, K. (2011). Virtual muscularity: A content analysis of male video game characters. *Body Image*, *8*(1), 43–51. https://doi.org/https://doi.org/10.1016/j.bodyim.2010.10.002

Matallaoui, A., Koivisto, J., Hamari, J., & Zarnekow, R. (2017). How effective is "exergamification"? A systematic review on the effectiveness of gamification features in exergames. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.

McArthur, V. (2018). Making mii: Studying the effects of methodological approaches and gaming contexts on avatar customization. *Behaviour & Information Technology*, 1–14. https://doi.org/10.1080/0144929X.2018.1526969

McEachan, R. R. C., Conner, M., Taylor, N. J., & Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the theory of planned behaviour: A meta-analysis. *Health Psychology Review*, *5*(2), 97–144.

McElreath, R. (2016a). *Rethinking: Statistical rethinking book package*.

McElreath, R. (2016b). *Statistical rethinking: A bayesian course with examples in r and stan* (Vol. 122). CRC Press.

McGrayne, S. B. (2011). *The theory that would not die: How bayes' rule cracked the enigma code, hunted down russian submarines, & emerged triumphant from two centuries of controversy*. Yale University Press.

McIntyre, K., Paolini, S., & Hewstone, M. (2016). Changing people's views of outgroups through individual-to-group generalisation: Meta-analytic reviews and theoretical considerations. *European Review of Social Psychology*, *27*(1), 63–115.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141.

Messinger, P. R., Ge, X., Stroulia, E., Lyons, K., Smirnov, K., & Bone, M. (2008). On the relationship between my avatar and myself. *Journal for Virtual Worlds Research*, *1*(2).

Messner, M. A. (1988). Sports and male domination: The female athlete as contested ideological terrain. *Sociology of Sport Journal*, *5*(3), 197–211.

Miller, W. C., & Miller, T. A. (2010). Attitudes of overweight and normal weight adults regarding exercise at a health club. *Journal of Nutrition Education and Behavior*, *42*(1), 2–9.

Minecraft's player count continues its upward trajectory. (n.d.).

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement. *Systematic Reviews*, *4*(1), 1.

Monaghan, L. F. (2005). Discussion piece: A critical take on the obesity debate. *Social Theory & Health*, *3*(4), 302–314.

More, K. R., Phillips, L. A., & Colman, M. H. E. (2019). Evaluating the potential roles of body dissatisfaction in exercise avoidance. *Body Image*, *28*, 110–114.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123.

Morling, B., & Lamoreaux, M. (2008). Measuring culture outside the head: A meta-analysis of individualism—collectivism in cultural products. *Personality and Social Psychology Review*, *12*(3), 199–221.

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., … others. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515.

Moskowitz, G. B., & Carter, D. (2018). Confirmation bias and the stereotype of the black athlete. *Psychology of Sport and Exercise*, *36*, 139–146.

Mulgrew, K. E., McCulloch, K., Farren, E., Prichard, I., & Lim, M. S. C. (2018). This girl can #jointhemovement: Effectiveness of physical functionality-focused campaigns for women's body satisfaction and exercise intent. *Body Image*, *24*, 26–35. https://doi.org/https://doi.org/10.1016/j.bodyim.2017.11.007

Murray, H. A. (1943). Thematic apperception test.

Myers, R. S., & Roth, D. L. (1997). Perceived benefits of and barriers to exercise and stage of exercise adoption in young adults. *Health Psychology*, *16*(3), 277.

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the foundations of positive psychology* (pp. 239–263). Springer.

Nakamura, L. (2013). *Cybertypes: Race, ethnicity, and identity on the internet*. Routledge.

Neumark-Sztainer, D., Paxton, S. J., Hannan, P. J., Haines, J., & Story, M. (2006). Does body satisfaction matter? Five-year longitudinal associations between body satisfaction and health behaviors in adolescent females and males. *Journal of Adolescent Health*, *39*(2), 244–251.

Neustaedter, C., & Fedorovskaya, E. (2009). Presenting identity in a virtual world through avatar appearances. In *Proceedings of graphics interface 2009* (pp. 183–190). Canadian Information Processing Society.

Nintendo. (2020a). Earnings release for the nine-month period ended december 2019. Retrieved from https://www.nintendo.co.jp/ir/pdf/2020/200130_3e.pdf

Nintendo. (2020b). Top selling title sales units. Retrieved from https://www.nintendo.co.jp/ir/en/finance/software/index.html

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250.

Nocon, M., Hiemann, T., Müller-Riemenschneider, F., Thalau, F., Roll, S., & Willich, S. N. (2008). Association of physical activity with all-cause and cardiovascular mortality: A systematic review and meta-analysis. *European Journal of Cardiovascular Prevention & Rehabilitation*, *15*(3), 239–246.

Noon, E. J., & Meier, A. (2019). Inspired by friends: Adolescents' network homophily moderates the relationship between social comparison, envy, and inspiration on instagram. *Cyberpsychology, Behavior, and Social Networking*, *22*(12), 787–793.

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*, 30–53.

Ntoumanis, N., Guerrero, M. D., Gadeke, C., & Thøgersen-Ntoumani, C. (2018). Do exerciser weight status and perceived motivation predict instructors'motivation and beliefs about the exerciser? A test of motivation contagion effects. *Body Image*, *26*, 10–18. https://doi.org/https://doi.org/10.1016/j.bodyim.2018.05.001

Obana, K., Hasegawa, D., & Sakuta, H. (2017). Change in subjective evaluation of weight by the proteus effect. In *International Conference on Human-Computer Interaction* (pp. 353–357). Springer.

O'Brien, K. S., Hunter, J. A., Halberstadt, J., & Anderson, J. (2007). Body image and explicit and implicit anti-fat attitudes: The mediating role of physical appearance comparisons. *Body Image*, *4*(3), 249–256.

Ogilvie, D. M. (1987). The undesired self: A neglected variable in personality research. *Journal of Personality and Social Psychology*, *52*(2), 379.

Operario, D., & Fiske, S. T. (2004). Stereotypes: Content, structures, processes, and context.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173.

O'reilly, M., & Parker, N. (2013). "Unsatisfactory saturation": A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research*, *13*(2), 190–197.

Paolini, S., & McIntyre, K. (2019). Bad is stronger than good for stigmatized, but not admired outgroups: Meta-analytical tests of intergroup valence asymmetry in individual-to-group generalization experiments. *Personality and Social Psychology Review*, *23*(1), 3–47.

Patomäki, H., & Wight, C. (2000). After postpositivism? The promises of critical realism. *International Studies Quarterly*, *44*(2), 213–237.

Paxton, S. J., Neumark-Sztainer, D., Hannan, P. J., & Eisenberg, M. E. (2006). Body dissatisfaction prospectively predicts depressive mood and low self-esteem in adolescent girls and boys. *Journal of Clinical Child and Adolescent Psychology*, *35*(4), 539–549.

Pazhoohi, F., Silva, C., Lamas, J., Mouta, S., Santos, J., & Arantes, J. (2018). The effect of height and shoulder-to-hip ratio on interpersonal space in virtual environment. *Psychological Research*. https://doi.org/10.1007/s00426-017-0968-1

Pearl, R. L., Dovidio, J., & Puhl, R. M. (2015a). Visual portrayals of obesity in health media: Promoting exercise without perpetuating weight bias. *Health Education Research*, *30*(4), 580–590.

Pearl, R. L., & Puhl, R. M. (2018). Weight bias internalization and health: A systematic review. *Obesity Reviews*, *19*(8), 1141–1163.

Pearl, R. L., Puhl, R. M., & Dovidio, J. F. (2015b). Differential effects of weight bias experiences and internalization on exercise among women with overweight and obesity. *Journal of Health Psychology*, *20*(12), 1626–1632.

Peeters, G. (1983). Relational and informational patterns in social cognition. *Current Issues in European Social Psychology*, *1*, 201–237.

Peng, W., Crouse, J. C., & Lin, J.-H. (2012). Using active video games for physical activity promotion. *Health Education & Behavior*, *40*(2), 171–192. https://doi.org/10.1177/1090198112444956

Peng, W., Lin, J.-H., & Crouse, J. (2011). Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. *Cyberpsychology, Behavior, and Social Networking*, *14*(11), 681–688.

Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 544.

Peña, J., Hancock, J. T., & Merola, N. A. (2009). The priming effects of avatars in virtual settings. *Communication Research*, *36*(6), 838–856.

Peña, J., Khan, S., & Alexopoulos, C. (2016). I am what i see: How avatar and opponent agent body size affects physical activity among men playing exergames. *Journal of Computer-Mediated Communication*, *21*(3), 195–209. https://doi.org/10.1111/jcc4.12151

Peña, J., & Kim, E. (2014). Increasing exergame physical activity through self and opponent avatar appearance. *Computers in Human Behavior*, *41*, 262–267.

Persson, S., Benn, Y., Dhingra, K., Clark-Carter, D., Owen, A. L., & Grogan, S. (2018). Appearance-based interventions to reduce uv exposure: A systematic review. *British Journal of Health Psychology*, *23*(2), 334–351.

Persson, S., Grogan, S., Dhingra, K., & Benn, Y. (2018). "It's bit of an eye opener"–a qualitative study of women's attitudes towards tanning, sun protection and a facial morphing intervention. *Psychology & Health*, *33*(3), 381–397.

Pila, E., Barlow, M. A., Wrosch, C., & Sabiston, C. M. (2016). Comparing the body to superior others: Associations with daily exercise and body evaluation in men and women. *Psychology of Sport and Exercise*, *27*, 120–127. https://doi.org/https://doi.org/10.1016/j.psychsport.2016.08.001

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, *64*(2), 191–199.

Popper, K. (1963). Science: Conjectures and refutations. *Conjectures and Refutations*, 43–77.

Pridgeon, L., & Grogan, S. (2012). Understanding exercise adherence and dropout: An interpretative phenomenological analysis of men and women's accounts of gym attendance and non-attendance. *Qualitative Research in Sport, Exercise and Health*, *4*(3), 382–399. https://doi.org/10.1080/2159676X.2012.712984

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting* *and Clinical Psychology*, *51*(3), 390.

Puhl, R. M., & Brownell, K. D. (2006). Confronting and coping with weight stigma: An investigation of overweight and obese adults. *Obesity*, *14*(10), 1802–1815.

Puhl, R. M., & Heuer, C. A. (2010). Obesity stigma: Important considerations for public health. *American Journal of Public Health*, *100*(6), 1019–1028.

Puhl, R. M., Latner, J., O'Brien, K., Luedicke, J., Daníelsdóttir, S., & Forhan, M. (2015). A multinational examination of weight bias: Predictors of anti-fat attitudes across four countries. *International Journal of Obesity*, *39*(7), 1166–1173.

Puhl, R. M., Peterson, J. L., DePierre, J. A., & Luedicke, J. (2013). Headless, hungry, and unhealthy: A video content analysis of obese persons portrayed in online news. *Journal of Health Communication*, *18*(6), 686–702.

Puhl, R., & Suh, Y. (2015). Health consequences of weight stigma: Implications for obesity prevention and treatment. *Current Obesity Reports*, *4*(2), 182–190. https://doi.org/10.1007/s13679-015-0153-z

Quinn, D. M., Kallen, R. W., & Cathey, C. (2006). Body on my mind: The lingering effect of state self-objectification. *Sex Roles*, *55*(11-12), 869–874.

Rahmani, E., & Boren, S. A. (2012). Videogames and health improvement: A literature review of randomized controlled trials. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, *1*(5), 331–341.

Rare. (2014). Kinect sports rivals.

Rasinaho, M., Hirvensalo, M., Leinonen, R., Lintunen, T., & Rantanen, T. (2007). Motives for and barriers to physical activity among older adults with mobility limitations. *Journal of Aging and Physical Activity*, *15*(1), 90–102.

Ratan, R. (2010). Self-presence, explicated. In *60th annual conference of the international communication association, singapore*.

Ratan, R., Beyea, D., Li, B. J., & Graciano, L. (2019). Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect. *Media Psychology*, 1–25.

Ratan, R., & Dawson, M. (2016). When mii is me: A psychophysiological examination of avatar self-relevance. *Communication Research*, *43*(8), 1065–1093.

Ratan, R., & Hasler, B. (2009). Self-presence standardized: Introducing the self-presence questionnaire (spq). In *Proceedings of the 12th Annual International Workshop on Presence*.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Re, A. C. D. (2013). *Compute.es: Compute effect sizes. R Package*. Retrieved from http://cran.r-project.org/web/packages/compute.es

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61.

Reeder, G. D., Pryor, J. B., & Wojciszke, B. (1992). Trait–behavior relations in social information processing.

Revelle, W. (2018). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rheu, M., Jang, Y., & Peng, W. (2019). Enhancing healthy behaviors through virtual self: A systematic review of health interventions using avatars. *Games for Health Journal*.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331–363.

Robertson, N., & Vohora, R. (2008). Fitness vs. Fatness: Implicit bias towards obesity among fitness professionals and regular exercisers. *Psychology of Sport and Exercise*, *9*(4), 547–557. https://doi.org/https://doi.org/10.1016/j.psychsport.2007.06.002

Robinson, B. “. E., Bacon, L. C., & O’reilly, J. (1993). Fat phobia: Measuring, understanding, and changing anti-fat attitudes. *International Journal of Eating Disorders*, *14*(4), 467–480.

Robinson, L., Prichard, I., Nikolaidis, A., Drummond, C., Drummond, M., & Tiggemann, M. (2017). Idealised media images: The effect of fitspiration imagery on body satisfaction and exercise behaviour. *Body Image*, *22*, 65–71. https://doi.org/https://doi.org/10.1016/j.bodyim.2017.06.001

Rooij, A. de, Land, S. van der, & Erp, S. van. (2017). The creative proteus effect: How self-similarity, embodiment, and priming of creative stereotypes with avatars influences creative ideation. In *Proceedings of the 2017 acm sigchi conference on creativity and cognition* (pp. 232–236). ACM.

Rosch, E. (1999). Principles of categorization. *Concepts: Core Readings*, *189*.

Ross, A. M., & Melzer, T. (2016). Beliefs as barriers to healthy eating and physical activity. *Australian Journal of Psychology*, *68*(4), 251–260. https://doi.org/10.1111/ajpy.12103

Royall, R. (2017). *Statistical evidence: A likelihood paradigm*. Routledge.

Rudd-Center. (2019). Media gallery. Retrieved from http://uconnruddcenter.org/media-gallery

Sah, Y. J., Ratan, R., Tsai, H.-Y. S., Peng, W., & Sarinopoulos, I. (2017). Are you what your avatar eats? Health-behavior effects of avatar-manifested self-concept. *Media Psychology*, *20*(4), 632–657.

Sailes, G. A. (1993). An investigation of campus stereotypes: The myth of black athletic superiority and the dumb jock stereotype. *Sociology of Sport Journal*, *10*(1), 88–97.

Samitz, G., Egger, M., & Zwahlen, M. (2011). Domains of physical activity and all-cause mortality: Systematic review and dose–response meta-analysis of cohort studies. *International Journal of Epidemiology*, *40*(5), 1382–1400.

Schaefer, L. M., Burke, N. L., Thompson, J. K., Dedrick, R. F., Heinberg, L. J., Calogero, R. M., … others. (2015). Development and validation of the sociocultural attitudes towards appearance questionnaire-4 (sataq-4). *Psychological Assessment*, *27*(1), 54.

Schaefer, L. M., Harriger, J. A., Heinberg, L. J., Soderberg, T., & Kevin Thompson, J. (2017). Development and validation of the sociocultural attitudes towards appearance questionnaire-4-revised (sataq-4R). *International Journal of Eating Disorders*, *50*(2), 104–117.

Schleicher, D. J., & McConnell, A. R. (2005). The complexity of self–complexity: An associated systems theory approach. *Social Cognition*, *23*(5), 387–416.

Schneider, M. C., & Bos, A. L. (2014). Measuring stereotypes of female politicians. *Political Psychology*, *35*(2), 245–266.

Schwartz, M. B., Vartanian, L. R., Nosek, B. A., & Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity*, *14*(3), 440–447.

Seacat, J. D., & Mickelson, K. D. (2009). Stereotype threat and the exercise/ dietary health intentions of overweight women. *Journal of Health Psychology*, *14*(4), 556–567. https://doi.org/10.1177/1359105309103575

Shaw, A. (2015). *Gaming at the edge: Sexuality and gender at the margins of gamer culture*. Univ Of Minnesota Press.

Shih, M., Wout, D. A., & Hambarchyan, M. (2015). Predicting performance outcomes from the manner of stereotype activation and stereotype content. *Asian American Journal of Psychology*, *6*(2), 117.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.

1177/0956797611417632

Simons, H. D., Bosworth, C., Fujita, S., & Jensen, M. (2007). The athlete stigma in higher education. *College Student Journal*, *41*(2), 251–274.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*(4), 689.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131.

Skowronski, J. J., & Carlston, D. E. (1992). Caught in the act: When impressions based on highly diagnostic behaviours are resistant to contradiction. *European Journal of Social Psychology*, *22*(5), 435–452.

Snow, J., & Mann, M. (2013). Qualtrics survey software: Handbook for research professionals. *Qualtrics Labs, Inc.*

Sofi, F., Valecchi, D., Bacci, D., Abbate, R., Gensini, G. F., Casini, A., & Macchi, C. (2011). Physical activity and risk of cognitive decline: A meta-analysis of prospective studies. *Journal of Internal Medicine*, *269*(1), 107–117.

Song, H., Kim, J., Kwon, R. J., & Jung, Y. (2013a). Anti-smoking educational game using avatars as visualized possible selves. *Computers in Human Behavior*, *29*(5), 2029–2036.

Song, H., Kim, J., & Lee, K. M. (2014). Virtual vs. Real body in exergames: Reducing social physique anxiety in exercise experiences. *Computers in Human Behavior*, *36*, 282–285. https://doi.org/http://dx.doi.org/10.1016/j.chb.2014.03.059

Song, H., Kim, J., Tenzek, K. E., & Lee, K. M. (2013b). The effects of competition and competitiveness upon intrinsic motivation in exergames. *Computers in*

Human Behavior, *29*(4), 1702–1708.

Stan Development Team. (2018). RStan: The R interface to Stan. Retrieved from http://mc-stan.org/

Statistica. (n.d.). South africa video games market value 2014-2023 published by christina gough, nov 7, 2019 the timeline presents the value of the video game market in south africa from 2014 to 2018, with a forecast until 2023. According to the calculations, the video game market in south africa will grow from 3.5 billion south african rand in 2018 to 5.44 billion in 2023. Value of the video games market in south africa from 2014 to 2023. Retrieved from https://www.statista.com/statistics/517310/south-africa-video-games-market-value/

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, *69*(5), 797.

Stephan, W. G., Ageyev, V., Stephan, C. W., Abalakina, M., Stefanenko, T., & Coates-Shrider, L. (1993). Measuring stereotypes: A comparison of methods using russian and american samples. *Social Psychology Quarterly*, 54–64.

Stephenson, N. (2014). *Snow crash*. Bragelonne.

Stern, R. M., Botto, R. W., & Herrick, C. D. (1972). Behavioral and physiological effects of false heart rate feedback: A replication and extension. *Psychophysiology*, *9*(1), 21–29.

Stone, A. R. (1991). Will the real body please stand up? *Cyberspace: First Steps*, 81–118.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, *77*(6), 1213.

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2015). Human height is positively related to interpersonal dominance in dyadic interactions. *PLoS One*, *10*(2), e0117860.

Sung, B., & Phau, I. (2019). When pride meets envy: Is social superiority portrayal in luxury advertising perceived as prestige or arrogance? *Psychology & Marketing*, *36*(2), 113–119.

Sween, J., Wallington, S. F., Sheppard, V., Taylor, T., Llanos, A. A., & Adams-Campbell, L. L. (2014). The role of exergaming in improving physical activity: A review. *Journal of Physical Activity and Health*, *11*(4), 864–870.

Sylvia, Z., King, T. K., & Morse, B. J. (2014). Virtual ideals: The effect of video game play on male body image. *Computers in Human Behavior*, *37*, 183–188.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (7th ed.). Allyn & Bacon/Pearson Education.

Talbot, C. V., Gavin, J., Van Steen, T., & Morey, Y. (2017). A content analysis of thinspiration, fitspiration, and bonespiration imagery on social media. *Journal of Eating Disorders*, *5*(1), 40.

Täuber, S., Gausel, N., & Flint, S. W. (2018). Weight bias internalization: The maladaptive effects of moral condemnation on intrinsic motivation. *Frontiers in Psychology*, *9*, 1836.

Team, J., & others. (2018). JASP (version 0.8. 6)[Computer software].

Tengler, C. D., & Jablin, F. M. (1983). Effects of question type, orientation, and sequencing in the employment screening interview. *Communications Monographs*, *50*(3), 245–263.

The chinese gamer. (n.d.).

Thompson, J. K., & Stice, E. (2001). Thin-ideal internalization: Mounting evidence for a new risk factor for body-image disturbance and eating pathology. *Current Directions in Psychological Science*, *10*(5), 181–183.

Thompson, V. A., & Pennycook, G. (2016). Base-rate neglect. In *Cognitive illusions* (pp. 54–71). Psychology Press.

Tiggemann, M., & Anesbury, T. (2000). Negative stereotyping of obesity in children: The role of controllability beliefs. *Journal of Applied Social Psychology*, *30*(9), 1977–1993.

Tiggemann, M., & Rothblum, E. D. (1988). Gender differences in social consequences of perceived overweight in the united states and australia. *Sex Roles*, *18*(1-2), 75–86.

Tiggemann, M., & Zaccardo, M. (2015). "Exercise to be fit, not skinny": The effect of fitspiration imagery on women's body image. *Body Image*, *15*, 61–67. https://doi.org/https://doi.org/10.1016/j.bodyim.2015.06.003

Townend, L. (2009). The moralizing of obesity: A new name for an old sin? *Critical Social Policy*, *29*(2), 171–190. https://doi.org/10.1177/0261018308101625

Tracy, J. L., & Prehn, C. (2012). Arrogant or self-confident? The use of contextual knowledge to differentiate hubristic and authentic pride from a single nonverbal expression. *Cognition & Emotion*, *26*(1), 14–24.

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: A tale of two facets. *Journal of Personality and Social Psychology*, *92*(3), 506.

Trepte, S., & Reinecke, L. (2010). Avatar creation and video game enjoyment: Effects of life-satisfaction, game competitiveness, and identification with the avatar. *Journal of Media Psychology*, *22*(4), 184;171;–184.

Turkay, S., & Kinzer, C. K. (2014). The effects of avatar-based customization on player identification. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *6*(1), 1–25.

Turkle, S. (1997). Multiple subjectivity and virtual community at the end of the freudian century. *Sociological Inquiry*, *67*(1), 72–84.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Ukiepedia. (2020, May). Player diversity & demographics. Retrieved from https://ukiepedia.ukie.org.uk/index.php/Player_Diversity_%26_Demographics

United States. Department of Health. (2008). *2008 physical activity guidelines for americans: Be active, healthy, and happy!* (Vol. 36). Government Printing Office.

Vainshelboim, B., Brennan, G. M., LoRusso, S., Fitzgerald, P., & Wisniewski, K. S. (2019). Sedentary behavior and physiological health determinants in male and female college students. *Physiology & Behavior*.

Valins, S. (1966). Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, *4*(4), 400.

Van der Eijk, C., & Rose, J. (2015). Risky business: Factor analysis of survey data–assessing the probability of incorrect dimensionalisation. *PloS One*, *10*(3), e0118900.

Van de Ven, N., Zeelenberg, M., & Pieters, R. (2012). Appraisal patterns of envy and related emotions. *Motivation and Emotion*, *36*(2), 195–204.

Vane, J. R. (1981). The thematic apperception test: A review. *Clinical Psychology Review*, *1*(3), 319–336.

Van Nguyen, H., Huang, H.-C., Wong, M.-K., Lu, J., Huang, W.-F., & Teng, C.-I. (2016). Double-edged sword: The effect of exergaming on other forms of exercise; a randomized controlled trial using the self-categorization theory. *Computers in Human Behavior*, *62*, 590–593.

Vartanian, L. R., & Novak, S. A. (2011). Internalized societal attitudes moderate the impact of weight stigma on avoidance of exercise. *Obesity*, *19*(4), 757–762.

Vartanian, L. R., & Shaprow, J. G. (2008). Effects of weight stigma on exercise motivation and behavior: A preliminary investigation among college-aged females. *Journal of Health Psychology*, *13*(1), 131–138.

Vehtari, A., Gabry, J., Yao, Y., & Gelman, A. (2018). Loo: Efficient leave-one-out cross-validation and waic for bayesian models. Retrieved from https://CRAN.

R-project.org/package=loo

Verhulst, A., Normand, J.-M., Lombart, C., Sugimoto, M., & Moreau, G. (2018). Influence of being embodied in an obese virtual body on shopping behavior and products perception in vr. *Frontiers in Robotics and AI*, *5*, 1–20.

Vicdan, H., & Ulusoy, E. (2008). Symbolic and experiential consumption of body in virtual worlds: From (dis) embodiment to symembodiment. *Journal for Virtual Worlds Research*, *1*(2).

Vorauer, J. D., Hunter, A., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology*, *78*(4), 690.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … others. (2018). Bayesian inference for psychology. Part i: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57.

Wallace, D. F. (2011). *The pale king*. Little, Brown.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*(5), 456–467.

Wasilenko, K. A., Kulik, J. A., & Wanic, R. A. (2007). Effects of social comparisons with peers on women's body satisfaction and exercise behavior. *International Journal of Eating Disorders*, *40*(8), 740–745. https://doi.org/10.1002/eat.20433

Wattanasoontorn, V., Boada, I., Garcia, R., & Sbert, M. (2013). Serious games for health. *Entertainment Computing*, *4*(4), 231–247.

Webb, T. L., & Sheeran, P. (n.d.). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychol Bull*, *132*(2), 249–268. https://doi.org/10.1037/0033-2909.132.2.249

Willig, C. (2013). *Introducing qualitative research in psychology*. McGraw-Hill Education (UK).

Wininger, S. R., & White, T. (2008). The dumb jock stereotype: To what extent do student-athletes feel the stereotype? *Journal for the Study of Sports and Athletes in Education*, *2*(2), 227–237.

Wininger, S. R., & White, T. A. (2015). An examination of the dumb jock stereotype in collegiate student-athletes: A comparison of student versus student-athlete perceptions. *Journal for the Study of Sports and Athletes in Education*, *9*(2), 75–85.

Wojciszke, B. (2005). Affective concomitants of information on morality and competence. *European Psychologist*, *10*(1), 60–70.

Wolin, K. Y., Yan, Y., Colditz, G. A., & Lee, I. (2009). Physical activity and colon cancer prevention: A meta-analysis. *British Journal of Cancer*, *100*(4), 611.

Wollersheim, D., Merkes, M., Shields, N., Liamputtong, P., Wallis, L., Reynolds, F., … others. (2010). Physical and psychosocial effects of wii video game use among older women. *International Journal of Emerging Technologies and Society*, *8*(2), 85–98.

Woodcock, J., Franco, O. H., Orsini, N., & Roberts, I. (2010). Non-vigorous physical activity and all-cause mortality: Systematic review and meta-analysis of cohort studies. *International Journal of Epidemiology*, *40*(1), 121–138.

World Health Organisation. (n.d.). Physical activity and adults. Retrieved from https://www.who.int/dietphysicalactivity/factsheet_adults/en/

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://github.com/rstudio/bookdown

Yang, G. S., Gibson, B., Lueke, A. K., Huesmann, L. R., & Bushman, B. J. (2014). Effects of avatar race in violent video games on racial attitudes and aggression. *Social Psychological and Personality Science*, *5*(6), 698–704.

Yee, N. (2007). *The proteus effect: Behavioral modification via transformations of digital self-representation* (PhD thesis). Stanford University.

Yee, N. (2014). *The proteus paradox: How online games and virtual worlds change us-and how they don't*. Yale University Press.

Yee, N., & Bailenson, J. N. (2009). The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology*, *12*(2), 195–209. https://doi.org/10.1080/15213260902849943

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, *36*(2), 285–312.

Yee, N., Ducheneaut, N., Yao, M., & Nelson, L. (2011). Do men heal more when in drag?: Conflicting identity cues between user and avatar. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 773–776). ACM.

Young adult educational and employment outcomes by family socioeconomic status. (n.d.). Retrieved from https://nces.ed.gov/programs/coe/indicator_tbe.asp

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*.