UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title

How do expertise and realism moderate the boundary between real and digital faces?

Permalink

https://escholarship.org/uc/item/8wn9p042

Author

Day, Jennifer Anne

Publication Date

2019

Copyright Information

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

HOW DO EXPERTISE AND REALISM MODERATE THE BOUNDARY BETWEEN REAL AND DIGITAL FACES?

A dissertation submitted in partial satisfaction of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Jennifer Day

June 2019

The Dissertation of Jennifer Day is
approved:

_____

Professor Nicolas Davidenko, chair

_____

Professor Travis Seymour

_____

Professor Katherine Isbister

_____

Lori Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Tables and Figures

Abstract

How do expertise and realism moderate the boundary between real and digital faces?

Jennifer Day

I explore a new approach to studying the development of face expertise by examining digital representations of faces. Typically digital faces have lower recognition scores than real faces, and I propose expertise and style as factors that moderate this boundary. Experiment 1 recruited participants who have played over 50 hours of The Elder Scrolls V: Skyrim as an expert population (experts, n=51) and compared their score on an upright and inverted face recognition task against participants who had not played Skyrim (novices, n=55). We also tested two different races of faces from Skyrim (Nord and Altmer). Participants performed significantly better on upright faces, t(104) =14.056, p <.0001 and Altmer faces, t(104) =-5.346, p < .001, and experts performed significantly better than novices, t(103) =2.664, p < 0.01. Participants then watched a video of gameplay with eye-tracking where participants' proportions of fixations on Skyrim faces were recorded. This face proportion measure did not significantly correlate with any other measure. Finally, participants completed a survey with questions about the number of hours played, video game habits, open-ended questions about experiences in Skyrim, and a character recognition task. Experiment 2 recruited a new population of novice participants (n=46) and compared

scores on an upright and inverted face recognition task for morphed photos

of human faces to 3 different styles of video game faces that ranged from

highly realistic (Monster Hunter: World), moderately realistic (Skyrim), to

highly stylized (Blade & Soul). Participants performed significantly better on

upright faces over inverted faces, $F(1, 45) = 13.06$, $p < .0001$ and on realistic

faces over stylized faces, $F(1, 45) = 54.657$, $p < .0001$. Participants

demonstrated a significantly larger upright advantage for stylized faces over

realistic faces $F(1, 45) = 11.97$, $p < .01$. Participants then completed a survey

with questions about video game habits, perceived task difficulty, and

looking strategies. The results for experiments 1 and 2 provide evidence to

suggest that both expertise and style can account for reduced performance in

digital faces, and open up questions about how these factors interact.

Acknowledgments

Despite the frequent and consistent use of the word "I" in this document, I would like to clarify that I did not do it alone. It would not have been possible without Hannah Hart-Pomerantz, Erika Lin, and Natalie Earl, who helped with data organization, stimuli creation, participant recruitment, analysis, and literature review. The analysis code for the eye-tracking experiment was developed by Matthew Jacobs. In addition, the members of the High Level Perception Lab were crucial for data collection.

Nicolas Davidenko was a huge contributor to this project, assisting with experiment design, coding, writing, and paper organization. He has been a thoughtful and dedicated mentor throughout this process. My committee contributed insightful feedback on my design and provided novel viewpoints and applications for the project that helped to make it more well-rounded as a whole. Douglas Bonett provided sage statistics advice. Fellow grad student Acacia Overoye proposed the idea for an other race effect with Skyrim faces at the beginning of my grad career.

Throughout this process I had encouragement from the important and wonderful people in my life; my family, friends, cohort, lab-mates, and partners. Their support was essential to the completion of this project.

To all of you, thank you.

CHAPTER I

Introduction

Perceptual expertise is the ability to distinguish and discriminate between exemplars of the same visual category (Scott, 2011). Humans typically show perceptual expertise for faces, which manifests in infant studies (Goren, Sarty, & Wu, 1975; Mondloch et al., 1999), neuroimaging (Kanwisher, McDermott, & Chun, 1997; Kanwisher, 2000), and adult studies (Yin, 1969; Ellis, 1975; Bruce & Young 1986; Tanaka & Farah, 1993; Richler, Cheung, & Gauthier, 2011). This expertise is flexible and can shift based on meaningful exposure to new faces. Changes in perceptual expertise can occur over a course of hours, as demonstrated with training studies (Gauthier & Tarr, 1997; Wong, Palmeri, and Gauthier, 2009), but also occur over longer periods of time, as in the *other-race effect* (Valentine & Bruce, 1986; Meissner & Brigham, 2001). Comparing expertise biases produced by short-term exposure to biases generated from a lifetime of exposure can give valuable insight into how faces are organized, encoded, and stored in the brain.

Research has examined not only how humans develop expertise for faces as an object class, but also how humans gain expertise for specific groups of faces over others (Bukach, Gauthier, & Tarr, 2006; McKone et al., 2012; Tanaka & Farah, 1993; Farah, Wilson, Drain, & Tanaka, 1998; Maurer, Le Grand, & Mondloch, 2002; Richler, Cheung, & Gauthier, 2011; Malpass & Kravitz, 1969; Meissner & Brigham, 2001). However, little is known about

how expertise develops with digital representations of faces such as avatars in video games, social media contexts, and other digitally rendered face stimuli. With the increasing ubiquity of digital representations of faces, it is important to study how expertise develops for them across both long and short-term exposure. Determining how people generate expertise for digital faces provides a novel and ecologically valid approach to examine how face expertise can change over time.

In Chapter 2, I will explore the literature around perceptual expertise for faces and how it develops in digital face representations. Perceptual expertise for faces is explored in depth in the following ways: the development of perceptual expertise, short-term perceptual expertise, long-term perceptual expertise, and the other-race effect. I then consider how perceptual expertise may extend to digital faces by considering the following: the differences between real faces and digital faces and developing expertise with video game characters.

CHAPTER II

Development of Perceptual Expertise

From the time they are born, infants show a preference for faces over other object categories (Fantz, 1963; Goren, Sarty, & Wu, 1975; Maurer & Young, 1983; Johnson et al., 1991). They also quickly develop a preference for *specific* faces, such as attractive faces (Slater et al., 1998), faces with direct gaze (Farroni et al., 2002), smiling faces (Farroni et al., 2007), and familiar identities such as their mother (Bushnell, 1989; Pascalis et al., 1995). This suggests infants have complex face representations (Quinn & Slater, 2003) and prioritize faces for social interaction (Johnson, 2011).

Johnson & Morton (1991) proposed a two-process model for the development of face perception that combines two theories: the tendency for infants to orient towards face-like objects (CONSPEC) and the resulting neural specialization for face recognition (CONLERN). They argued that subcortical processing was responsible for guiding face perception in infants from birth to 4 weeks, based on low spatial frequency information (Le Doux, 1996). Neuroimaging studies have identified the specific areas involved in this pathway: the superior colliculus, pulvinar, and amygdala (George et al., 2001; Johnson, 2005; Kleinhans et al., 2008). Evidence from ERP and MEG studies show that this pathway is significantly faster than the adult face-coding pathway (Bailey et al., 2005). The infant pathway likely guides the preference for face-like objects and lays the foundation for the

development of a social brain network. Evidence from PET and ERP studies suggest that cortical brain regions that are activated when adults process faces (such as the fusiform face area, or FFA, and occipital face area, or OFA) are activated by face stimuli early in postnatal development (Johnson, 2011). This evidence supports the theory of interactive specialization that posits that specialized brain regions develop from more broadly functional areas (Johnson, 2001; Johnson, 2011).

Over the early years of life, face perception seems to shift. In the first few hours of life, infants are largely relying on the external features of a face for recognition (Perrett, 2010). If a mother changes her hairstyle, for example, visual recognition will fail. This pattern is consistent across studies; infants' face processing first relies on external features and then shifts to rely on internal features (Maurer & Salapatek, 1976; Heith, 1977; Pascalis, 1995; Bartrip, 2001; Blass & Camp, 2004). This shift is accompanied by an increased ability to process individual identities. The more familiar a face is, the more likely infants are to rely on internal features to identify them (Campbell et al., 1999; Clutterbuck & Johnston, 2004; Ellis et al., 1979). This corresponds with an increased benefit for infants recognizing faces that hold direct gaze (Farroni et al., 2007; Hood et al., 2003) and increased ability for recognition of the eye region (Ge et al., 2008; Goldstein & Mackenberg, 1966; Hay & Cox, 2000).

As infants gain more expertise with recognizing individual identities, their exposure to specific groups of faces begins to shape their expertise. Infants as young as 6 months show increased recognition ability for faces of the race they have the most contact with, which manifests as the "other-race effect" (Hayden et al., 2007). Kelly, Quinn, and Slater et al. (2007) demonstrated Caucasian infants were equally able to recognize Caucasian, Chinese, Middle Eastern, and African faces at 3 months, but stopped demonstrating the ability to recognize Middle Eastern and African faces at 6 months, and only demonstrated recognition of Caucasian faces at 9 months. In another experiment, Kelly, Liu, and Lee et al. (2009) demonstrated that Chinese infants showed a similar pattern, and could only recognize Chinese faces at 9 months. However, the other-race effect is significantly more fragile in infants than in adults, as even brief experience with other-race faces can eliminate the effect entirely in infants (Sangrigoli & de Schonen, 2004).

Experience not only plays a key role in early perceptual development, but it also shapes expertise throughout the lifespan (Bukach, Gauthier, & Tarr, 2006; Tanaka & Pierce, 2009; McKone et al., 2012). By the time we are adults, we have already developed expertise in distinguishing a number of visual categories such as faces, bodies, and animals (Scott, 2011). This expertise manifests both neurally and behaviorally (Gauthier et al., 1998; Scott et al., 2008; Scott et al., 2006; McKone, Kanwisher, & Duchaine, 2007; Wong et al., 2009). Typically, expertise is studied one of two ways: (1)

training novices to become experts with a visual object class or (2) comparing expert populations to non-expert populations.

     Short-term perceptual expertise. Training studies are employed to track the development of expertise with novel objects over a relatively short period of time. Gauthier and Tarr (1997) first employed this technique to train people to become experts with 'greebles', cylindrical objects with four protruding appendages. Each individual greeble belonged one of two 'genders' and one of five 'families'. Training occurred over multiple sessions, and it typically took 10 hours until participants were considered experts (Bukach, Gauthier, & Tarr, 2006). Compared to novices, greeble experts demonstrated faster reaction times for a greeble identification task (Gauthier & Tarr, 1997), increased reliance on configural information (Gauthier et al., 1998), increased activation in face-selective regions of the cortex such as the FFA and OFA (Gauthier, 1999; Gauthier & Tarr, 2002), and increased response in electrophysical face processing indices such as the N170 ERP component (Rossion et al., 2002).

     There are a few key aspects to training studies that give insight into how expertise is developed. One key aspect of developing expertise is that it is important to categorize the object (or face) on a deeper level. Rosch et al. (1976) defined the levels of abstraction as superordinate, basic, and subordinate. There are many ways to classify a specific face: human (superordinate level), female (basic level), Caucasian female (subordinate

level), or Meryl Streep (individual level; Rosch et al.,1976). Participants must learn to classify objects at either the subordinate or individual level to gain expertise with an object class (Tanaka & Taylor, 1991; Tanaka, 2001; Scott et al., 2006; Scott et al., 2008; Nishimura & Maurer, 2008; Wong, Palmeri, & Gauthier, 2009). For example, Wong, Palmeri, and Gauthier (2009) trained two separate groups on a novel object class called Ziggerins. One group learned to categorize Ziggerins into basic level categories, whereas the other learned to categorize them at the individual (or subordinate) level. Only participants who were trained at the subordinate level showed an increase in performance on a sequential matching task and a reduction in reaction time on a triplet recognition task when processing the Ziggerins.

Another important aspect of developing expertise is receiving feedback. Mere exposure is not enough to develop perceptual expertise (Scott et al., 2008). Expertise can only be created through meaningful exposure to an object class that provides learning at different levels and is accompanied by feedback. For example, Krigolson (2009) trained participants to discriminate between two different categories of blob stimuli. Participants would select a category and receive either positive or negative feedback. Some of the blob stimuli were morphs between the two object categories and participants were provided with random (either negative or positive) feedback. Their results showed learning and expertise only for the

non-morph blob stimuli, implying that consistent feedback is important for the development of perceptual expertise.

Although the feedback in this example is specific to the goals of the experiment, feedback is common in real life scenarios. In the real world, feedback for learning to individuate faces can come in many forms, for example being able to tell the difference between your father and uncle, learning someone's name for the first (or second) time, or mistaking a stranger for someone you know. The negative social consequences of these errors can be thought as a type of feedback.

Training studies also provide valuable insight into the time course of the development of perceptual expertise. Both studies of Ziggerins and greebles reported around 10 hours of training time (Gauthier & Tarr, 1997; Wong, Palmeri, and Gauthier, 2009). Despite the relatively low amount of training time, it was sufficient to elicit neural and behavioral changes. Further insight into expertise can be gleaned through comparing novices to experts that have developed expertise over years of experience.

Long-term perceptual expertise. Comparing experts to non-experts can also provide valuable insight into how long-term expertise manifests behaviorally and neurally. As discussed earlier, experts with object classes process the objects more configurally and at deeper levels of abstraction. These behavioral effects have been demonstrated robustly when looking at face perception (Gauthier et al., 1998; Wong et al., 2009; Tanaka, 2001), but

there are also studies that show these effects when looking at other objects of expertise such as dogs, birds, and cars (Diamond & Carey, 1986; Tanaka & Taylor, 1991; Gauthier, Skudlarski, Gore, & Anderson, 2000).

Expertise in object classes (like faces) allows observers to distinguish identities by noticing subtle differences in the spatial relations of features (Tanaka & Farah, 1993; Farah, Wilson, Drain, & Tanaka, 1998; Maurer, Le Grand, & Mondloch, 2002; Tanaka, & Sengco, 1997; Richler, Cheung, & Gauthier, 2011; Young, Hellawell, & Hay, 2013). This is called second-order configural processing and expert observers can use this configural information for fast and accurate recognition. In behavioral experiments, the occurrence of configural processing is tested against situations where this processing is disrupted.

It is well-established that inversion will disrupt the configural processing of faces; a face stimulus that is flipped upside-down will significantly reduce a person's ability to recognize it (Yin, 1969; Rhodes, Brake, & Atkinson, 1993; Farah, Tanaka, & Drain, 1995; Freire, Lee, & Symons, 2000; Leder & Bruce, 2000; Yovel & Kanwisher, 2005). This effect is also found with other object classes; for example, Diamond & Carey (1986) demonstrated that dog experts are impaired in dog identification tasks when the dogs are inverted. Similar inversion effects have been found for car experts (Rossion & Curran, 2010) and novices when recognizing human bodies (Reed et al., 2003). However, the inversion effect is more robustly

demonstrated for faces than for any other object class. This can likely be attributed to the high-level of expertise most people generate through a lifetime of experience with faces, as even experts of other object classes will still likely also show face expertise (Carey, 1992; Carey & Diamond, 1977; Tanaka & Gauthier, 1997).

Expertise with object classes also manifests behaviorally as deeper levels of abstraction. Rosch et al. (1976) demonstrated that basic-level object categories have the highest cue validity and therefore are most likely to be used by the general populace. Experts, however, tend to categorize objects of expertise at the subordinate (or individual) level. Tanaka and Taylor (1991) found that experts were more likely than novices to name objects by their subordinate level and suffered no reaction time differences; for example dog experts were more likely to name the dog breed (e.g. "Beagle") when shown an image of a dog, and were just as fast to classify it by breed as to name the object class ("dog"). Tanaka (2001) demonstrated that non-experts consistently named faces at the subordinate level (e.g. "Michelle Obama" or "a black woman") and non-face objects at the basic level. Participants were able to categorize faces at the subordinate level with the same speed they could categorize them at the superordinate level but showed slower reaction times when asked to name images of dogs at the subordinate level.

In addition to behavioral evidence, neuroimaging studies also show evidence of long term expertise causing changes in the brain. Specific areas in

the brain dedicated to object recognition show increased activity when processing faces (Kanwisher, McDermott, & Chun 1997; Haxby, Hoffman, & Gobbini, 2000; Grill-Spector, Knouf, & Kanwisher, 2004). In addition, some studies have shown similar patterns of activation for other objects of expertise (Tarr & Gauthier, 2000; Gauthier et al., 2000; Xu, 2005).

Kanwisher, McDermott, and Chun (1997) used fMRI to define a region of interest in the fusiform gyrus that showed increased activation when participants viewed faces over other objects. The FFA (fusiform face area) has since been the focus of an ongoing debate about where and how object expertise manifests in the brain.

Tarr and Gauthier (2000) argue that the FFA is domain-general and activates for within-category discrimination when any object of expertise is processed at the subordinate level. They provide evidence for this by demonstrating increased activation in the FFA and LOC (Lateral Occipital Complex) to a trained object class (greebles). Gauthier et al. (2000) measured fMRI responses to faces, familiar objects, birds, and cars in bird and car experts and found increased FFA and OFA activity when experts were tested on the location of their objects of expertise. Xu (2005) argued that this expertise effect could be due to experts processing these objects like faces. Participants were tested on modified stimuli of birds and cars that reduced how 'face-like' they appeared and found an expertise effect in the right FFA. However, these studies only showed activation when the task

tested participants on the spatial location of the stimuli and showed no correlations when participants were tested on object identity (Moore et al., 2006).

Kanwisher and Yovel (2006) posit that the FFA is domain-specific and activates only for detection and identification of faces and face-like objects. Grill-Spector, Knouf, and Kanwisher (2004) provided evidence for this theory by comparing FFA activation for car experts viewing cars to typical FFA activation for faces. Subordinate identification of cars was correlated more highly with other regions of the ventral occipitotemporal cortex than the FFA. Moore et al. (2006) trained participants (for less than 10 hours) on a non-face-like object class and found an increase in activity for the FFA and LOC during the cue period, but not after a delay. Yue, Tjan, and Biederman (2006) trained participants (for 8 hours) to become experts with blobs that required discriminating smoothly varying surfaces similar to the low-level information in faces. Blob experts did not show any increased activation of the FFA but did show increased activation in the LOC. These studies suggest that expertise effects for other objects could possibly be found in the LOC under specific testing conditions, but only face-like objects consistently show increased activation in the FFA.

Long term perceptual expertise manifests both behaviorally and neurally. Typically object expertise is thought of in terms of increased ability to recognize individual objects (such as faces) at the subordinate level, but

expertise can also vary *within* an object class. For example, increased

experience with a specific race, age, or gender of faces can manifest

behaviorally and neurally as expertise for those specific types of faces over

others.

The other-race Effect

The other-race effect is a salient example of variable expertise within

an object class. It is characterized by the tendency for people to more easily

recognize and distinguish faces of the race that they are most familiar with

(Malpass & Kravitz, 1969; Levin, 1996; Sporer, 2001; Meissner & Brigham,

2001). Meissner and Brigham (2001) published a meta-analysis on research

that was done between 1970 and 2000 on the other-race effect. Overall,

participants were 1.4 times more likely to correctly identify a previously

viewed own-race face and 1.56 times more likely to falsely identify a novel

other-race face. The effect generally replicated across studies and different

memory tasks but was inconsistent across racial groups. Increasing

interracial contact reduced the other-race effect, but also slightly decreased

sensitivity to same-race faces. Sensitivity here is defined as the likelihood to

notice differences between similar exemplars. There are three main theories

researchers utilize to explain the other-race effect; contact hypothesis,

race-coding hypothesis, and multidimensional face-space model.

Contact hypothesis posits that the other-race effect is caused by

increased exposure to the race of expertise over other races (Brigham &

Malpass, 1985, Chiroro & Valentine 1995; Furl, Phillips, & O'Toole, 2002; Hancock & Rhodes, 2008). Contact hypothesis suggests that the less exposure an observer has with a particular race of faces, the lower the accuracy in face recognition tasks with that race. Typically, the other-race effect is more robust in largely homogeneous populations, and less evident in multiracial populations (Meissner & Brigham, 2001; Chiroro & Valentine, 1995; Bar-Haim et al., 2006). The experience with identifying faces of one's own race (or the race of expertise) also determines which face recognition strategies an individual uses. Part-whole, inversion, and composite face tasks have demonstrated that identifying an own race face (or face of expertise) not only increases the amount someone processes the face as a whole (holistic processing), but also sensitivity to featural and relational differences (Murray et al., 2003; Tanaka et al., 2003; Michel et al., 2006; Mondloch et al., 2010).

The race-coding hypothesis addresses the different levels of processing between own race and other race faces. The race-coding hypothesis postulates that the decreased recognition for other race faces is a result of coding race-specifying information over individuating information (Rhodes et al., 1989; Rhodes et al., 2009). Rhodes et al., (2009) demonstrated an increased other-race effect when participants were asked to code race-specifying information for both other and same-race faces. Levin (1996) demonstrated that participants were faster at categorizing other race faces

but slower at individuating other race faces. This theory relates to the idea that with increased expertise, people will become better at subordinate processing (Rosch et al., 1976; Tanaka & Taylor, 1991; Tanaka, 2001). This provides additional evidence that own race faces have increased expertise but also presents the idea that other race faces are processed at a more intermediate (or basic) level. Chance and Goldstein (1981) asked white participants to describe their first reactions to white, black, and Japanese faces. The responses were assigned a score on a scale of 0 (very shallow) to 100 (very deep). White faces showed significantly higher ratings (M = 78.4) than black (M = 61.6) and Japanese (M = 57.3) faces, indicating a deeper level of processing for own-race faces. Maclin and Malpass (2001) demonstrated that external cues for race cued participants to race code racially ambiguous face stimuli as own- or other-race, which manifested as higher individuation for faces participants coded as their own race.

The multidimensional face-space model suggests that the difference in processing of own race and other race faces is due to how tightly clustered the representations are in face space (Byatt & Rhodes, 2004; Papesh & Goldinger, 2010; Valentine, 1991; 2016). Face space proposes that faces are represented as encoded points in a multidimensional space. A face's location in this multidimensional space provides an appropriate parallel to the mental representation of a face. Faces that are more visually similar are grouped closer together in this space, and the average face (or norm) is situated in the

15

direct center. The points in a person's face space represent all faces a person has been exposed to over a lifetime, which means each person will have a differently populated face space. For example, if a person is exposed to a high number of faces of their own race, it is likely that the average face will tend towards that race. Valentine defines these individual spaces as a person's implicit knowledge of faces. Races with which a participant is more familiar have better representation and are less tightly clustered in face space. Meaningful exposure to other race faces can reduce the other-race effect.

A complete understanding of the other-race effect requires integration of the contact hypothesis, race-coding hypotheses, and the multidimensional face-space model. The categorization-individuation model suggests that categorization (basic level of processing) and individuation (subordinate level of processing) and are two separate processes that play different roles in how the other-race effect manifests (Hugenberg et al., 2010). For own race faces (or races one has experience with) there is a tendency to attend to identity-diagnostic characteristics such as configural information. This helps explain the increase in holistic processing when identifying own race faces (Tanaka et al., 2003; Michel et al., 2006; Mondloch et al., 2010). For other race faces (or races one has less experience with) there is a tendency to attend to category diagnostic features such as skin tone. This helps explain the faster categorization times for other race faces

and the more shallow levels of processing (Levin, 1996; Chance & Goldstein, 1981; Tanaka, 2001).

The categorization-individuation model is also supported by evidence from experiments that attempt to *reduce* the other-race effect via training. Research suggests that meaningful exposure to other-race faces can reduce the other-race effect (Goldstein & Chance, 1985; McKone et al., 2007; Hugenberg, Miller, & Claypool, 2007; Rhodes et al., 2009; Pauker et al., 2009; Young, Bernstein, & Hugenberg, 2010). Meaningful exposure can be thought of as training participants to attend to identity-diagnostic (subordinate) information. Goldstein and Chance (1985) demonstrated such an effect of training; white participants were trained to distinguish between pairs of Japanese faces (for around 4 hours, over a couple of weeks). This training mitigated the other-race effect for Japanese faces, and the effect persisted when participants were tested 5 months later. McKone et al. (2007) demonstrated the emergence of normal accuracy and holistic processing for other race faces after familiarization (for an hour) with specific other race identities. The other-race effect can even be reduced by encouraging participants to individuate other race identities within the span of an experiment (Hugenberg et al., 2007; Rhodes et al., 2009; Pauker et al., 2009; Young et al., 2010). For example, Hugenberg, Miller, and Claypool (2007) warned participants about the other-race effect before testing and asked

them to focus on individuating other race faces. The external motivation

showed an improvement in recognizing other race faces.

Evidence strongly suggests that expertise for faces (and faces of

specific races) increases with individuation. Social or attentional cues that

require individuals to attend to identity-diagnostic information encourage a

deeper level of learning and experience with individual identities within

specific groups helps solidify this expertise. Does the type of expertise that

people demonstrate with humans transfer to other face like objects? The next

chapter explores the idea that expertise could transfer to digital

representations of faces in the context of repeated and consistent exposure

to video game faces in ecologically similar scenarios.

Chapter III

Differences between real and digital faces

Digital (or computer-generated) faces are being used more often as ease of access increases (e.g. Leopold et al., 2001; Todorov, Said, Engell, & Oosterhof, 2008; Papesh & Goldinger, 2010). Digital faces allow for manipulation and standardization of face stimuli, which is vital for studies that are trying to pinpoint specific differences between faces such as in adaptation studies (Leopold et al., 2001), social evaluation (Todorov, Said, Engell, & Oosterhof, 2008), and even the other-race effect (Papesh & Goldinger, 2010). However there has been research to suggest that digital faces differ from photographs of faces in face recognition accuracy (Green, MacDorman, Ho, & Vasudevan, 2008; Crookes et al., 2015; Balas and Pacella, 2015) and neural responses (Wheatley, Weinberg, Looser, Moran, & Hajcak, 2011, Balas & Koldewyn 2013; Schindler, Zell, Botsch, & Kissler, 2017).

Green, MacDorman, Ho, and Vasudevan (2008) first demonstrated that participants experienced increased sensitivity to changes in face shape for face photographs over computer-generated faces. Crookes et al. (2015) found that face photographs were recognized more accurately than computer-generated stimuli based on those photographs, and computer-generated faces based on real identities were recognized more accurately than computer-generated identities. Balas and Pacella (2015) found that performance on an old/new and a match-to-sample identity

recognition task was significantly better for real faces over computer-generated faces.

Wheatley, Weinberg, Looser, Moran, and Hajcak (2011) demonstrated that while both human and artificial faces showed early response in the N170 and VPP (vertex positive potential), only human faces continued to show activity beyond that. Balas and Koldewyn (2013) showed sensitivity to animacy in real faces present in the P100 and category differentiation at the N170 that depended on animacy. They suggested that responses in the N170 component reflected sensitivity to changes in faces that people have the most experience with, which in this case was natural (or real) faces over artificial faces. A study by Schindler, Zell, Botsch, and Kissler (2017) demonstrated a stronger electroencephalogram (EEG) response for real faces over computer-generated faces for the N170, EPN (early posterior negativity), and LPP (late positive potential). Real faces also activated different regions than computer-generated faces; real faces showed increased activation in middle and inferior occipital regions, whereas computer-generated faces showed highest activation levels in the right inferior occipital gyrus (Schindler, Zell, Botsch, and Kissler, 2017).

One theory for why people show reduced face recognition abilities for digital faces over real faces is the difference in perceptual expertise (Kätsyri, 2018). Using a face memory recognition task, Kätsyri (2018) demonstrated a higher false alarm rate (and higher response bias) for digital faces over real

faces. These results resemble the other-race effect in that other-race faces typically elicit higher false alarm rates (Meissner and Brigham, 2001). Because of reduced expertise with digital faces, observers may be less well tuned to the variation shown in computer-generated faces (Crookes et al., 2015).

Crookes et al. (2015) suggest three possibilities for the reduced ability to discriminate digital face identities; (1) digital faces contain less information to discriminate between identities and are overall more similar to each other than real faces, (2) human face processing abilities are less tuned to digital faces than to real faces, due to lack of exposure and experience with digital faces, and (3) due to the artificial nature of digital faces, they are considered out-group faces and are processed similarly to other-race faces.

Balas and Pacella (2015) suggest that experience with digital faces could modulate the boundary between own-group and out-group faces, similar to how experience with other race faces can reduce the other-race effect. They suggest that individual differences in exposure to digital faces in movies or games could account for the differences in performance for digital and real faces. Balas and Pacella (2015) also discuss the possibility that different strategies are used when discriminating between digital faces that are novel and digital faces with which a person is familiar.

Developing expertise with digital representations of faces.

As previously discussed, meaningful exposure to and practice with a group of faces generates expertise effects, which can be demonstrated by increased sensitivity and recognition (Malpass & Kravitz, 1969; Levin, 1996; Sporer, 2001; Meissner & Brigham, 2001). As digital representations of faces increase in ubiquity, it is possible that people are not only generating expertise with digital faces overall but also with specific styles of digital representations of faces. For example, consider a person's first exposure to a new digital context, such as a new online video game. When first presented with a certain stylistic representation of faces, it may be difficult to tell two avatars apart from one another. However, continued exposure and a need to individuate characters or avatars could generate expertise with that specific stylistic representation.

As it becomes increasingly important to discriminate one person's avatar from another, people may become more attuned to the subtle differences between them. Importance of character recognition ties back into the ideas of individuation and feedback. In some games, quests are difficult if you don't recognize the characters involved; for example, it is difficult to deliver fire salts to Balimund if you cannot remember what he looks like. A narrative can also become difficult to follow if you cannot recognize character faces; for example, if you don't recognize Delphine's face, you might become confused when she sheds her disguise to help negotiate the end of a civil war. In many roleplaying games, individual characters have names,

identities, and stories. Without identity recognition, this information becomes difficult to track and understand.

Research over the past decade has covered in detail how people perceive and relate to their own digital representations in video games; researchers study how avatar representations affect self-concept (Nowak et al., 2008; Manning, 2009; Scarborough & Bailenson, 2014; Banks, 2016), social interactions (Yee, Bailenson, Urbanek, Chang, & Merget, 2007; Ratan, 2010; Martey, 2011; Van der Heide, 2013), and user experience (Isbister, 2016; Guegan & Moliner, 2015; Williams, D., Kennedy, 2011; Kang, Watt, & Isbister, 2006; Tychsen et al., 2008). These studies provide strong evidence that the perception of the digital self is deeply rooted in how people experience the digital world. However, little has been studied about how players' relationships and experience with these digital representations create perceptual expertise. Testing expertise with digital face representations in video game players is a novel approach that provides a high level of ecological validity. Compared to lab-training studies that typically employ 10 or fewer hours of training to generate expertise, video game players generate this expertise over hundreds of hours of gameplay over the course of months or years.

As discussed in an earlier section (see The other-race effect), expertise effects are generated not through mere exposure, but through meaningful practice of distinguishing between identities (Meissner, 2001)

and receiving feedback. An experimental paradigm that utilizes video game character faces leverages existing expertise in a generation that has grown up with these faces and developed expertise, sometimes from a very young age. For digital faces that are closer to real faces (highly detailed), users may be able to draw from prior knowledge to assist with individuation. But differentiating between identities in simple or low-detail representations may require new expertise.

Gomez, Barnett, and Grill-Spector (2019) recently demonstrated that participants who played Pokémon in childhood showed Pokémon specific activation in the OTC region of the brain. This suggests that video games that require players to individuate exemplars as a mechanism of gameplay provide a similar enough context to how incidental learning occurs with objects of interest, such as faces, birds, dogs, or cars. The experience causes expertise to develop for the video game exemplars and manifests as a specific neural representation. Video games that have gameplay that requires individuation of digital face representations may be generating expertise with that specific style of representation.

I hypothesize that face expertise effects manifest with digital representations of faces in video game contexts where character recognition is important to the game narrative. I conducted a preliminary online questionnaire (n = 85) about the recognition of NPCs (non-player characters) in the video game, The Elder Scrolls V: Skyrim (Bethesda Game Studios,

2011). Participants who had played Skyrim were recruited through the subreddit r/skyrim but were not prescreened for expertise or number of hours played. Participants were asked how important they considered NPC recognition on a scale of 1 (not at all important) to 5 (extremely important). The mean score across 85 participants was 3.33, with 84.7% of participants reporting that they considered NPC recognition as moderately important or higher. Participants were also tested on their ability to recognize faces of specific characters from the game by name, with an average accuracy score of 73.8%. A log transform of the number of hours participants (n=80) self-reported spent playing the game was positively correlated with accuracy in this task ($r(79) = .59$, $p < .001$). Self-reported time spent playing the game ranged from 18 to 4000 hours, two to three orders of magnitude higher than the typical time course of training studies (10 hours, see Gauthier and Tarr, 1997). Further, 52.94% of players self-reported that they look at the character's face when engaging in dialogue. These data suggest NPCs are important enough for players to need individuation, game experts can recognize in-game characters, and many players look at faces when engaging in dialogue with an NPC. This ties into the idea that it is important to recognize characters from Skyrim to help track narratives, quests, and relationships.

In digital contexts where it is important to recognize identities, users may form expertise effects for the specific style of face representation they

are exposed to. Expertise with a style of face representation could lead to increased identity recognition and reliance on configural information with that style of faces. Here we can think of a manifestation of the other-race effect as the "other style effect". Even within a certain style of digital face representations, there may be categories (such as race or gender) that someone encounters more often than others. For example, if a role-playing game player encounters a specific race more often in the game, they may develop an other-race effect, even with fantasy races.

If the other-race effect transfers to digital game characters, it can be studied using a similar experimental paradigm. For example, people who play Skyrim may be better at distinguishing identities of *Nord* faces (a race of humans that have a high level of exposure to) than *Altmer* faces (a race of elves that have a low presence in the game).

Experiment 1 investigates how expertise with digital faces can manifest behaviorally. Specifically, experienced players of video games may develop expertise within that style of faces that manifests similarly to the other-race effect. I hypothesize that experienced players of video games will generate general expertise for the avatars of that game, which will manifest as increased recognition and a stronger upright advantage for those avatar faces when compared to novices. I also predict that experienced game players will also generate specific expertise for more frequently encountered races within that avatar style, whereas novices will not.

Experiment 2 investigates to what extent expertise with real faces

may be recruited for perceiving novel digital faces. As digital faces deviate

more away from realism and towards more stylized representations of faces,

people may start to perceive these faces as out-group faces. I hypothesize

that the boundary between realistic and stylized faces manifests similarly to

the boundary between own-group and out-group faces; novices (participants

with no experience with the specific stimuli sets) will perform better on

more realistic representations over more stylized representations.

The following experiments aim to define modulating factors for the

boundary between real and digital representations of faces. The first

experiment investigates how this boundary is modulated by expertise, and

the second experiment investigates how this boundary is modulated by

realism and style.

Chapter IV

Experiment 1: Testing Expertise and Other-Race Effects in

Skyrim Character Faces

Experiment 1 used an upright and inverted face recognition task to test if participants of different levels of experience with the video game The Elder Scrolls V: Skyrim show expertise and other-race effects for faces sourced from the game. Experiment 1 is broken down into three parts: 1a) upright and inverted face recognition task, 1b) eye tracking study, and 1c) survey. First, I expected expert players to show increased recognition performance for Skyrim avatars compared to novices. Second, I predict a pattern similar to the other-race effect to emerge in experts only: increased recognition and higher upright advantage for avatar faces within the high-exposure racial group compared to faces in the low-exposure racial group. This effect should be strongest for experts and weakest for non-experts and may be mediated by the number of hours played, number of face-directed fixations when viewing gameplay, and self-reported importance of character recognition.

Method

Participants. Two groups of participants were recruited at different levels of expertise: non-experts (55) and experts (57). Experts were defined as participants with 50+ hours of gameplay in The Elder Scrolls V: Skyrim. 6

participants were excluded from the expert group because they did not reach the 50-hour requirement, which brought the total number of experts to 51.

Sample size was determined with a power analysis based on the mean performance on upright Nord faces with novices in a pilot study that compared Nord faces to another race of Skyrim faces (n = 48, M = .647, SD = .149) and an assumption that experts will show better performance on upright Nord faces by a factor equivalent to the other-race effect in novices (.08).

Recruitment. Novice participants were recruited primarily through the SONA system at the University of California, Santa Cruz, and were given class credit in exchange for their time. Expert participants were recruited through the SONA system at the University of California, Santa Cruz and outside recruitment. Outside recruitment consisted of posting in online groups (such as the University of California, Santa Cruz student Facebook group), posting to the University of California, Santa Cruz subreddit, making an announcement in an introductory game design class, and posting flyers in public spaces (such as the bus stops on University of California, Santa Cruz campus). Participants recruited outside of the SONA pool were given a $10 Amazon gift card for their time. There was no significant difference in performance scores between experts who were paid and experts who were unpaid, $t(49) = 1.582$, $p > .05$.

Procedure. Participants came to the lab and completed a face recognition task (1A), an eye tracking task (1B), and a survey (1C) on laboratory computers. Each participant used only their initials and date to identify themselves on the face recognition task, eye-tracking task, and the survey.

Procedure for 1A. The behavioral part of the experiment is an upright and inverted face recognition task based on Valentine (1991) presented in Matlab. In each trial, a target avatar face was displayed to the participants in one of two views (either front view or ¾ view), followed by a mask, followed by the target face and two distractor faces at the other view (see Figure 1). The switching of views is included to increase the difficulty and encourage holistic processing (processing the face as a whole rather than by a specific feature). Half of the trials were done with upright faces and half were done with inverted faces. Participants were asked to select the target face by pressing a corresponding key, and reaction time and accuracy were recorded. The faces were displayed on an 8° x 8° square region on a monitor (screen resolution: 1920 x 1080), in this region, the faces subtended approximately 2° x 3°.
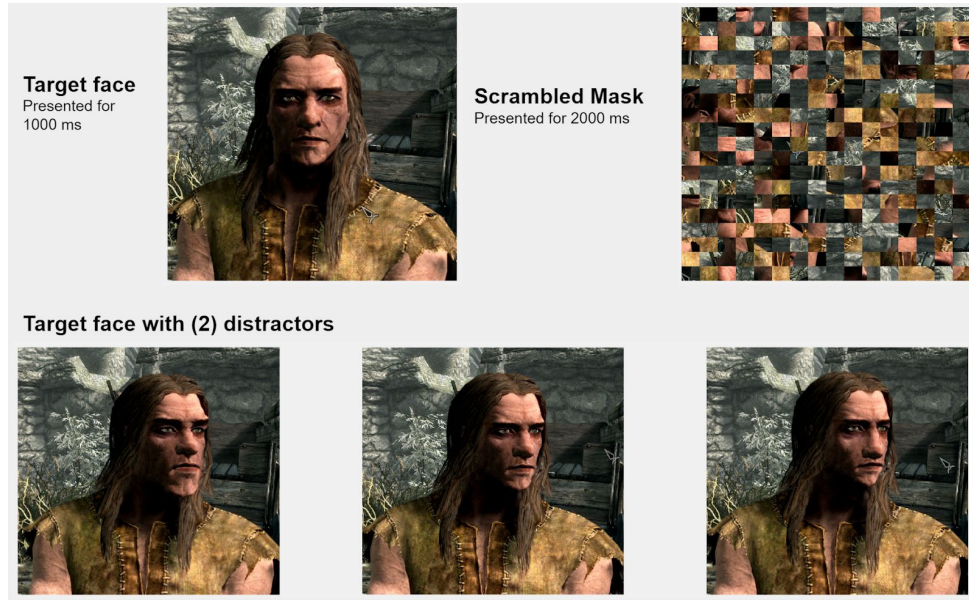
Figure 1: Example of procedure for Part 1A of Experiment 1. Character faces generated from the video game The Elder Scrolls V: Skyrim (Bethesda Game Studios, 2011).

Measures for 1A. There were three independent variables for 1A: expertise, orientation, and race. Expertise is a between-subjects variable with two levels (novices and experts). Orientation is a within-subjects variable with two levels (upright and inverted). Race is a within-subjects variable with two levels (Nord and Altmer faces). The dependent variables for 1A are two behavioral measures: reaction time and accuracy.

Stimuli for 1A. The stimuli were avatar faces created in the popular role-playing video game The Elder Scrolls V: Skyrim. This game was selected for three reasons: 1) the avatars in this game span multiple real and fantasy races, 2) the avatar creation system allows for easy capturing and editing of digital face stimuli, and 3) the game was rated as the most popular

roleplaying game in a poll of the online community at University of California, Santa Cruz. The stimuli were composed of 7 different identities of each avatar race, and three variants of each identity. Screenshots of each of the 3 identity variants were taken at two views. In each case, two of the variants were used as distractors, and one was used as a target.

A set of identities was created for each of 2 races; the Nords and the Altmer (for examples see Figure 2). The Nords are a caucasian-like race and very commonly encountered throughout gameplay, whereas the Altmer are a fantasy elf race that are very rarely encountered throughout gameplay. I hypothesize that experienced players will show expertise with a race that is more commonly encountered (the Nords) in a specific video game (The Elder Scrolls V: Skyrim) that will manifest as higher target recognition over another race that is less commonly encountered (the Altmer).

The face stimuli were constructed with the default character creator in The Elder Scrolls V: Skyrim. The game was run on ultra-high graphics settings (screen resolution: 1920 x 1080) to ensure the clearest images. Each identity was based on a different male preset for each race. Preset identities are constructed by game developers to provide additional starting points for character creation. Presets are a good example of a range of identities available to a player and are typically available for selection and modification. For each preset, 3 versions were constructed by adjusting the character creation sliders; 1) variant 1 moved the slider left for each feature

32

in the face, eyes, brow, and mouth category, 2) variant 2 moved the slider to

the right for each feature in the face, eyes, brow, and mouth category and 3)

variant 3 moved the slider to the direct center for each feature in the face,

eyes, brow, and mouth category. Screenshots of each variant were taken with

the face turned towards the front and ¾ to the right for a total of 6 images for

each identity.

To control for possible differences in variability between races, a

pairwise image variance analysis was conducted using the SSIM (Structural

Similarity Index) function in Matlab (Wang, Bovik, Sheikh, & Simoncelli,

2004). The average SSIM score for the Nord faces is 0.73 and the average

SSIM score for the Altmer faces is 0.75, and the difference between the two is

non-significant; $t(12) = -1.082$, p = 0.3.

Figure 2: A) Examples of Nord faces. B) Examples of Altmer faces.

Faces from the video game The Elder Scrolls V: Skyrim (Bethesda Game

Studios, 2011).

Procedure for 1B. 1B was an eye-tracking study run with a remotely

mounted eye-tracker (GazePoint GP3, 60Hz). All participants were run on the

eye-tracking portion of the experiment. Participants were calibrated for the

eye tracker and reminded to keep their eyes open and stay still throughout

the duration of the experiment. Participants were given headphones and

asked to watch a short 4-minute video while resting their head on a chin rest.

The video was a walkthrough of a town in The Elder Scrolls V: Skyrim.

Characters (20 different characters total) in the game speak to the

participant and to each other. While watching the video, the duration and

number of face-directed fixations were recorded.

Measures for 1B. 1B measured duration and number of face-directed

fixations per participant while watching the video. The data were exported as

a CSV file that contained XY coordinates for fixation location and frame

number (approximately 15,000 16.7-ms frames per participant). The total

number of frames was compressed to match the number of frames in the eye

tracking video (total frames: 748). For each participant, the XY coordinates of

each compressed frame were compared against the XY coordinates on a

template for the eye-tracking video annotated for where faces occurred in

the eye tracking video. This template was constructed at 3 frames per second

by annotating the XY coordinates of a circle around every face in a scene of

the video. A buffer variable was included to allow for an increase in the

circle's size around each face to account for possible errors with the

eye-tracking data (up to 1 degree). This allowed a metric to be computed that

determines the number of frames a participant fixated on a face by

approximating if a fixation was directed towards a face in the scene. I divided

the number of trials that a fixation was coded as face-directed by all fixations (during scenes with faces present) to compute the total proportion of fixations spent directed at a face.

Procedure for 1C. 1C was a survey run with Google Forms after parts 1A and 1B of the experiment were completed. All participants were asked questions on the following topics: the number of hours spent playing video games per week, age of first playing video games, self-identification of gamer stereotype, character recognition habits, and game familiarity. Expert participants were asked questions on the following topics: the number of hours played, recency of play, the importance of character recognition, level and type of interaction, and personal significance of narrative and gameplay. For the full survey see Appendix A. Expert participants were also asked to participate in an identity recognition task of characters in The Elder Scrolls V: Skyrim, which was run with Qualtrics.

Measures for 1C. The survey measured the number of hours played, recency of play, the importance of character recognition, level and type of interaction, the personal significance of narrative and gameplay, video game experience, and looking strategies. Experts also completed an identity recognition task where overall accuracy was recorded.

Hypotheses. I expected an effect of expertise; Skyrim experts should show increased recognition, lower reaction times, and more holistic processing (defined as a larger inversion effect) of avatar faces than

non-experts. I expected an interaction between avatar race and expertise; Skyrim experts should show increased recognition, lower reaction times, and increased holistic processing for faces of the races they have had the most experience with. I expected a main effect of orientation; faces that are presented upright should have higher accuracy and lower reaction times than faces shown inverted. I expected an interaction between orientation and expertise; participants with more expertise should show an increased difference between upright and inverted scores. I expected a 3-way interaction; Skyrim experts will show a larger disadvantage for inverted Nord faces over inverted Altmer faces; but Novices will not show this effect.

For non-experts, I expected that there will be an advantage in recognition performance, reaction times, and holistic processing forNord identities over Altmer identities (but the main effect of race should be smaller for Novices than Experts). I expected an effect of orientation for Novices (but the main effect of orientation should be smaller for Novices than Experts). For Skyrim experts, I expected that participants will show higher performance, lower reaction times, and increased holistic processing in the face recognition task for the commonly encountered race (the Nord identities) than the less commonly encountered race (the Altmer identities). This will manifest in the following ways: a significant increase in performance and lower reaction times for Nord identities over Altmer identities when stimuli are presented upright (main effect of race for upright

orientations), no difference in performance and reaction times for Nord identities compared to Altmer identities when stimuli are presented inverted (no main effect of race for inverted orientations), and a larger difference in performance and reaction times for upright vs. inverted Nord identities compared to upright vs. inverted Altmer identities (interaction effect of race and orientation). A higher number of face fixations, self-reported number of hours, and self-reported importance of character recognition should predict higher overall accuracy and lower reaction times in the face recognition task and a larger effect of orientation and avatar race.

I expected the proportion of face-fixations to positively correlate with the following behavioral measures from 1A; overall performance, upright advantage for performance, performance on Nord faces. I expected the proportion of face-fixations to negatively correlate with the following behavioral measures from 1A; overall reaction time and upright advantage for reaction time. I expected the proportion of face-fixations to positively correlate with the following survey measures from 1C; self-report of interaction frequency, self-reported importance of character recognition, and self-report of interaction frequency with Skyrim characters (for expert participants).

I expected average accuracy to be positively correlated with the following survey measures: reported starting age for playing video games, self-report of the importance of character recognition, self-report of

37

interaction frequency, number of hours played in Skyrim, self-report of looking strategy, and character recognition score. Because participants may be more likely to interact with characters if they prioritize identity recognition or relationships I expected self-report of interaction frequency and self-report of the importance of character recognition to be positively correlated.

Results

Results for 1A. Performance in the face recognition task was calculated as proportion correct by averaging scores for participants based on the within-subjects factors orientation and race of face (1 is correct, 0 is incorrect). Reaction times were collected for trials with correct answers and averaged across participant. Upright advantage scores were calculated by averaging the difference between accuracy for upright and inverted faces. The standard tested was p < .05 and the pairwise comparisons were adjusted for Least Significant Difference.

A three-way MANOVA with two repeated measures was conducted to measure effects of expertise (between subjects: novices, experts), race (within-subjects: Nord, Altmer), orientation (within-subjects: upright, inverted), and possible interactions between them on performance and reaction time (see Table 1). The alpha was at $p < .05$. There was a significant effect of expertise, Wilks' Lambda = .93, $F(2, 102) = 3.82$, $p < 0.05$. There was a significant effect of race, Wilks' Lambda = .768, $F(2, 102) = 15.44$, $p < 0.001$

(although it was in the opposite direction as predicted). There was a significant effect of orientation, Wilks' Lambda = .332, $F(2, 102) = 102.59$, $p <$ 0.001. There were no significant two-way interactions. There was a significant three-way interaction between expertise, race, and orientation, Wilks' Lambda = .942, $F(2, 102) = 3.139$, $p < 0.05$.

Table 1. Multivariate Effects

Variable(s)	Wilks' Lambda	F	df	Error df	Sig.
Expertise	0.93	3.818[1]	2	102	0.025
Race	0.768	15.441[1]	2	102	0
Race * Expertise	0.973	1.434[1]	2	102	0.243
Orientation	0.332	102.593[1]	2	102	0
Orientation * Expertise	0.97	1.575[1]	2	102	0.212
Race * Orientation	0.969	1.606[1]	2	102	0.206
Race * Orientation * Expertise	0.942	3.139[1]	2	102	0.048

Note. [1]Exact Statistic

Expertise. There was a significant effect of expertise on performance at the $p < .05$ level, $F(1, 103) = 7.098$, $p < .001$. Pairwise comparison of accuracy for expert and novice participants indicated that the average performance score for expert participants (M=0.59, SD=.013) was significantly higher than novice participants (M=0.54, SD=.013), $t(103)$ =2.664, $p = 0.009$ (see Figure 3). On average, expert participants scored .05 higher on accuracy than novices; 95% confidence interval for difference [.013, .087]. There was no significant effect of expertise on reaction time.
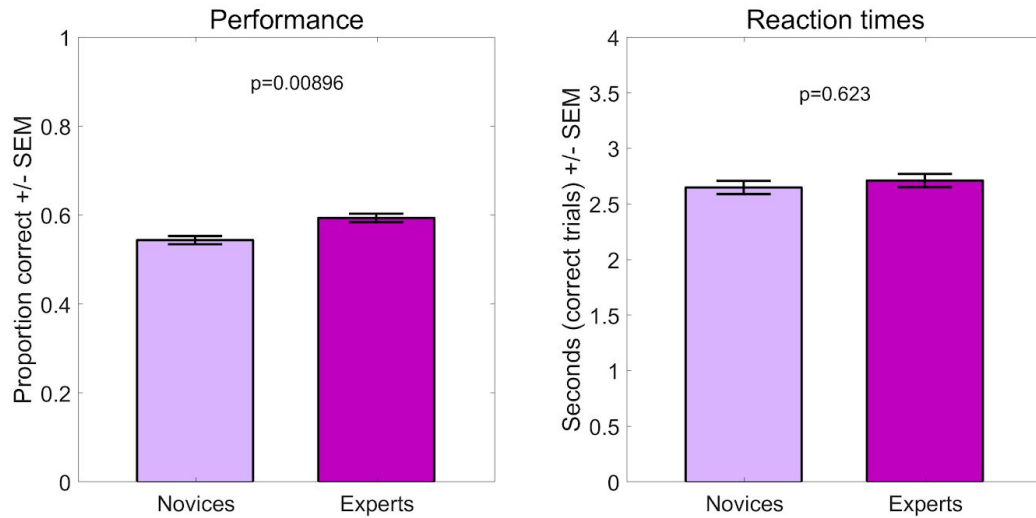
Figure 3: Average accuracy and reaction times for novice and expert participants.

Race. There was a significant effect of race on performance at the $p <$ .05 level, $F(1, 103) = 28.33$, $p < .001$. Pairwise comparison of accuracy for the Nord and Altmer conditions indicated that the average performance score for Nord faces (M=0.531, SD=.011) was significantly *lower* than the average performance score for Altmer faces (M=0.607, SD=.012), $t(104) =-5.346$, $p <$ .001, counter to my prediction. On average, participants scored .076 higher on accuracy for Altmer faces over Nord faces; 95% confidence interval for difference [.048, .104] (see Figure 4).

There was a significant effect of race on reaction time at the p<.05 level, F(1, 103) = 5.49, p < .05. Pairwise comparison of reaction time scores for the Nord and Altmer conditions indicated that the average reaction time for Nord faces (M=0.531, SD=.011) was significantly higher than the average reaction time for Altmer faces (M=0.607, SD=.012), $t(104) =2.219$, $p = 0.029$.

On average, participants responded .144 seconds faster for Altmer faces over Nord faces; 95% confidence interval for difference [.022, .266] (see Figure 4).
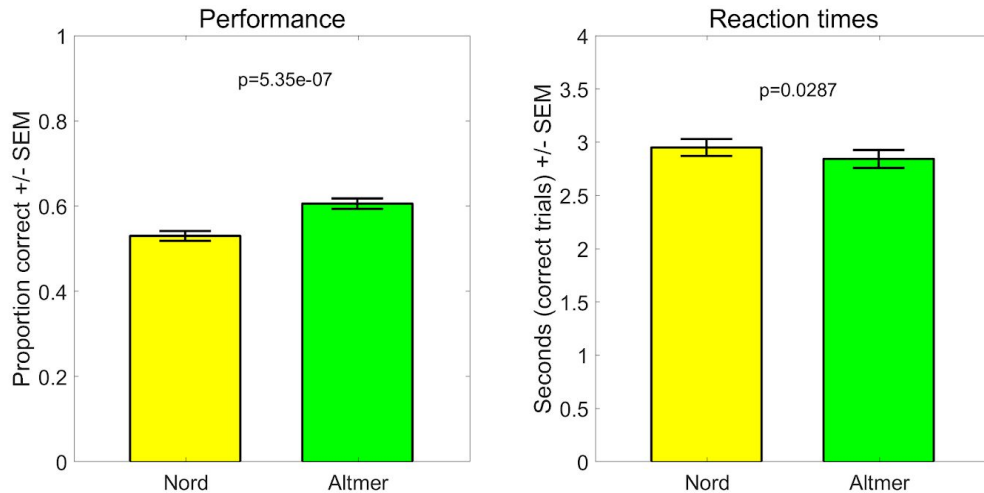
Figure 4: Average accuracy and reaction times for Nord and Altmer identities.

Orientation. There was a significant effect of orientation on performance at the *p* < .05 level, *F*(1, 103) = 201.98, *p* < .001. Pairwise comparison of accuracy for the upright and inverted conditions indicated that the average performance score for upright faces (M=0.67, SD=.012) was significantly higher than the average performance score for inverted faces (M=0.47, SD=.012), *t*(104) =14.056, *p* <.0001 (see Figure 5). On average, participants scored 0.199 higher on accuracy for upright faces over inverted faces; 95% confidence interval for difference [.171, .266].

There was a significant effect of orientation on reaction time at the *p* < .05 level, *F*(1, 103) = 40.230, *p* < .001. Pairwise comparison of reaction time scores for the upright and inverted conditions indicated that the average reaction time for upright faces (M=2.61, SD=.067) was significantly lower

41

than the average reaction time for inverted faces (M=3.10, SD=.099), $t(104)$ =-6.37, $p < .001$ (see Figure 5). On average, participants responded .492 seconds faster for upright faces over inverted faces; 95% confidence interval for difference [.338, .646].
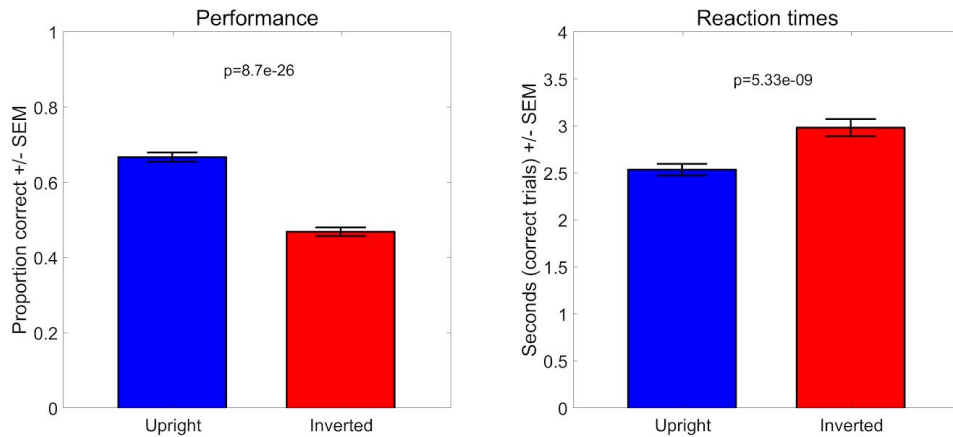
Figure 5: Average accuracy and reaction times for upright and inverted faces.

accuracy for upright and inverted faces were moderately positively correlated, $r(104) = 0.301$, $p = .002$ (see Figure 6). Reaction times for upright and inverted faces were moderately positively correlated, $r(104) = 0.644$, $p < .0001$ (see Figure 6). When comparing correlations separately for experts and novices, there is no significant correlation between upright and inverted faces for experts $r(50) = 0.241$, $p > .05$. However, there is a positive correlation for novices, $r(54) = 0.328$, $p < .05$.
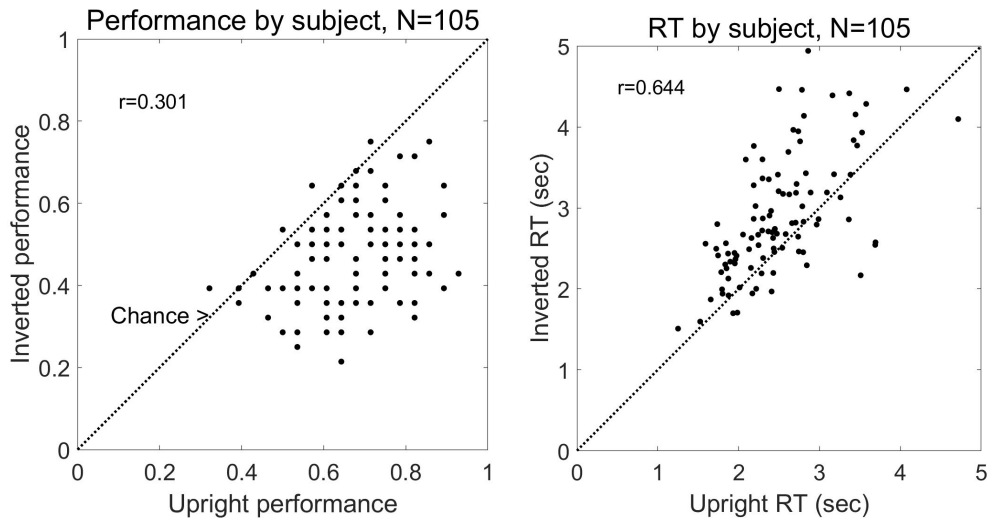
Figure 6: Scatter plots of average accuracy for upright and inverted faces and reaction times for upright and inverted faces.

Interactions. Interactions were calculated between expertise and race, expertise and orientation, race and orientation, and expertise, race, and orientation (see Table 2). There were no significant two-way interactions for performance: the interaction between expertise and race was non-significant, $F(1, 103) = .016$, $p = 0.898$, the interaction between expertise and orientation was non-significant, $F(1, 103) = 2.707$, $p = 0.103$, and the interaction between orientation and race was non-significant, $F(1, 103) = 3.105$, $p = 0.081$. There was no significant three-way interaction for performance: the interaction between orientation, expertise, and race was non-significant, $F(1, 103) = .376$, $p = 0.541$.

There were also no significant two-way interactions for reaction time: the interaction between expertise and race was non-significant, $F(1, 103) = 2.889$, $p = 0.092$, the interaction between expertise and orientation was

43

non-significant, $F(1, 103) = 1.3$, $p = 0.257$, and the interaction between orientation and race was non-significant, $F(1, 103) = .494$, $p = 0.484$. There was a significant interaction between expertise, race, and orientation for reaction time at the $p < .05$ level, $F(1, 103) = 6.322$, $p < .05$ (see Figure 7). This interaction suggests that experts were significantly faster for upright Nord faces over inverted Nord faces and that this difference was significantly larger than the difference in reaction times for upright and inverted Altmer faces. This difference was not found for novices: reaction times on Nord faces were not significantly larger than Altmer faces.

Table 2. Tests of Within-Subjects Contrasts

Variable(s)	Interaction	Sum of Squares	Mean Square	F	Sig.
Performance	Race * Expertise	0	0	0.016	0.898
	Orientation * Expertise	0.056	0.056	2.707	0.103
	Race * Orientation	0.053	0.053	3.105	0.081
	Race * Orientation * Expertise	0.006	0.006	0.376	0.541
Reaction time	Race * Expertise	1.143	1.143	2.889	0.092
	Orientation * Expertise	0.82	0.82	1.3	0.257
	Race * Orientation	0.224	0.224	0.494	0.484
	Race * Orientation * Expertise	2.872	2.872	6.322	0.013
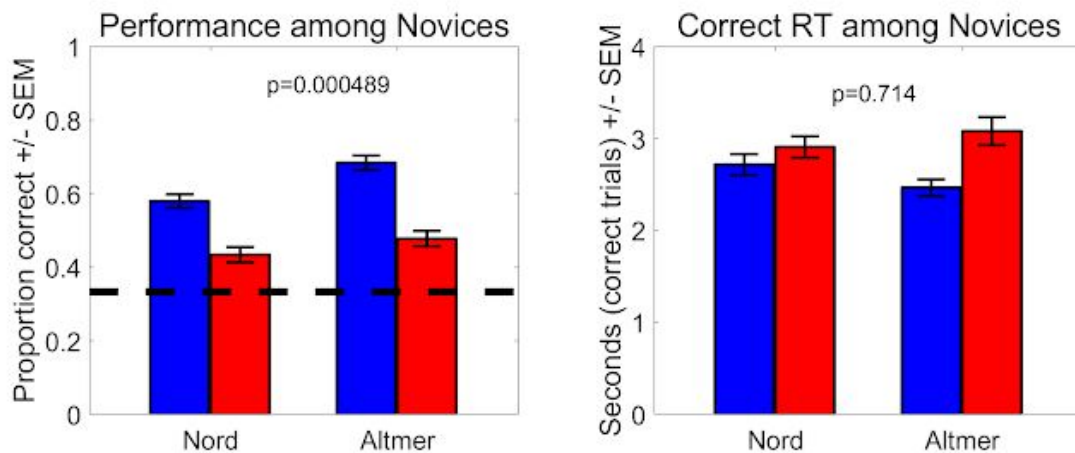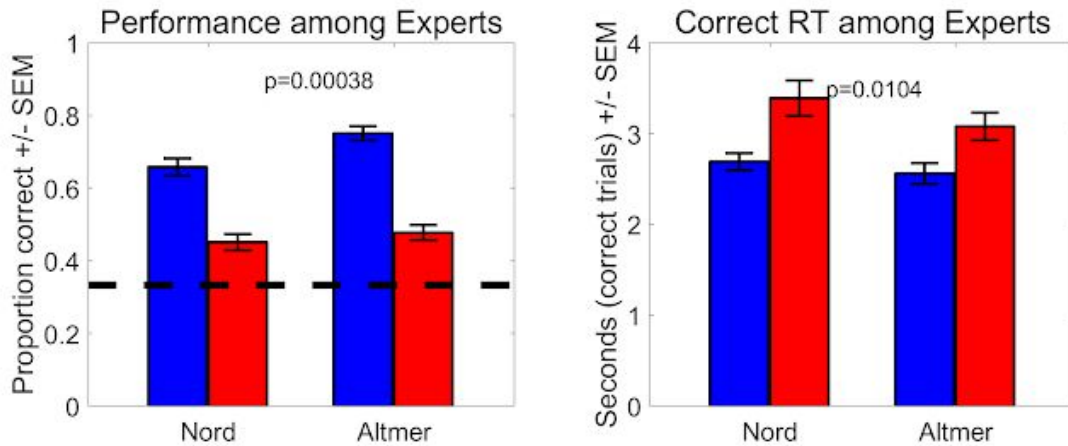
Figure 7: Average accuracy and reaction times for upright (blue) and inverted (red) faces, plotted by race of face (Nord and Altmer) and separated by the expertise of participants.

Discussion for 1A. There were significant main effects for each independent variable, and a significant three-way interaction. In this section, I discuss how these results relate to our hypotheses and address possible explanations.

Orientation. Participants were overall significantly more accurate and faster with upright faces over inverted faces. This corresponds with my hypothesis and the greater body of literature about the inversion effect. This suggests that Skyrim faces are processed as faces and are recruiting expertise mechanisms for real faces (similar to other schematic or representative faces). Orientation did not show any significant interaction effects with another single variable, which is discussed in the following sections.

45

Expertise. Expert participants scored over three standard deviations higher on accuracy than novices. This aligns with the hypothesis that experts have generated general expertise with faces from Skyrim, which shows in the overall performance increase over novices. Typically expertise effects are demonstrated with higher speed and accuracy but our data showed experts were not significantly faster or slower than novices. This suggests that experts are not simply showing an increase in performance because they are taking longer.

The expertise effect did not interact with orientation. I hypothesized that experts would show a larger upright advantage over novices, as this is a common behavioral manifestation for expertise. The upright advantage for experts (0.222) was greater than the upright advantage for novices (0.176) but the difference was non-significant, $p = 0.103$. The non-significant difference could indicate that 50 hours is not enough to generate expertise that manifests an upright advantage. This can be explored by testing the difference in upright advantage for two new expertise groups split at the median (250) number of hours. A two-sample t-Test revealed a non-significant difference in upright advantage for accuracy between the two expert groups, $t(49) = 0.05$, $p = 0.499$ and a non-significant difference in upright advantage for reaction times between the two expert groups, $t(49) = -0.002$, $p = 0.479$.

Another possible way to explore this is to compare experts with more than 250 hours to novices. However, a two-sample t-Test revealed a non-significant difference in upright advantage for accuracy between the experts (with 250+ hours) and novices, $t(77) = -1.16$, $p = 0.249$, and a non-significant difference in upright advantage for reaction times between the experts (with 250+ hours) and novices, $t(77) = .806$, $p =.423$. There was no significant differences in upright advantage between groups of participants, which could suggest more participants were needed to detect a difference.

Skyrim is filled with unique characters that a player can interact with, but many players report spending time out in the world exploring (n = 12, 23.5%). Participants who report enjoying activities that focus on exploring may show a reduced upright advantage compared to those who focus on character interaction (when controlling for the number of hours played). I calculated two survey scores based on self-reported activities that players enjoyed; exploration and character interaction.

I computed a two-stage hierarchical multiple regression for upright advantage in accuracy. The number of hours played was entered at stage one of the regression as a control, and the survey scores (exploration and character interaction) were entered at stage two. The hierarchical multiple regression revealed that at Stage one Number of hours, did not significantly contribute to the regression model, $F(1,49) = .046$, $p= .831$). At stage 2,

exploration and character interaction did not significantly contribute to the regression model, $F(1,47)$ = .240, p=.787). I also computed a two-stage hierarchical multiple regression for upright advantage in reaction time. The number of hours played was entered at stage one of the regression as a control, and the survey scores (exploration and character interaction) were entered at stage two. The hierarchical multiple regression revealed that at Stage one Number of hours, did not significantly contribute to the regression model, $F(1,49)$ = .636, p= .429). At stage 2, exploration and character interaction did not significantly contribute to the regression model, $F(1,47)$ = .838, p=.439).

The expertise effect also did not interact with race; which suggests that experts are not generating greater expertise with a specific subset of faces. I hypothesized that experts would be better with Nord faces over Altmer faces because Nords are one of the most common race in Skyrim and Altmers are one of the least common. There are two possibilities that I did not find an interaction with expertise and race; (1) the expertise generated for Skyrim faces is not deep or nuanced enough to show different effects for different subsets of this category and/or (2) players do not interact with Nord faces significantly more than Altmer faces.

I explored the first possibility by comparing the differences between Nord and Altmer faces for experts who have over 250 hours to experts who have 250 hours and less. I computed an Altmer advantage metric by finding

the difference in accuracy between Nord and Altmer faces for each participant. A two-sample t-Test revealed a non-significant difference in Altmer advantage between the two expert groups, $t(49) = 1.676$, $p = 0.192$.

The second possibility is more difficult to investigate as it is difficult to track the races of the characters our expert participants have interacted with most. Analyzing the race proportions of characters that the average player reports interacting with can give some insight into this question; there are approximately 1100 characters in Skyrim, Nord characters represent approximately 43.21% of these characters, compared to Altmer characters, which represent approximately 5.07% (Skyrim: Characters by Race, n.d.). In addition, an analysis was done to compare the race of an expert's favorite character (Nord, Altmer, or other) to the differences in Altmer advantage. No participants reported an Altmer character as their favorite character, 17 participants reported a Nord character as their favorite character, and 31 participants reported another race as their favorite character. A two-sample t-Test revealed a non-significant difference in Altmer advantage between the two groups, $t(46) = -1.168$, $p = 0.125$.

Race. Participants were overall significantly more accurate and faster with Altmer faces over Nord faces. I hypothesized that Nord faces would show this benefit across participant groups because of their more typical appearance and because they are more commonly encountered in Skyrim (for experts). Despite controlling for the similarity between Nord faces and

49

Altmer (by comparing the SSIM scores), I believe the benefit in recognition for Altmer faces is caused by the exaggerated features typical to the visualization of the race. Altmer faces typically have more pronounced and angular chins and noses than Nord faces (see Figure 8) and participants may be attending to these more distinct features to facilitate recognition strategies. The difference in scores between races might be accounted for in the future by asking participants to report their looking strategies.

Figure 8. Examples of Altmer faces; note the pronounced and angular chins and noses.

Another possible explanation for the benefit for recognition for Altmer faces is that the SSIM scores were not able to fully reflect the differences in facial similarity. It is possible that the Altmer distractors were more dissimilar to each other; this can be accounted for in future studies by performing a more specific similarity analysis that focuses on facial features and utilizing pilot participants for perceptual similarity ratings.

There was no interaction with race and expertise, which (as discussed above) could suggest that experts are not generating expertise with a specific

subset of faces. There is also the possibility that the performance benefit for Altmer faces outweighs any significant benefit an expert may have for Nord faces.

There was no interaction between race and orientation. I hypothesized that Nord faces would elicit a larger upright advantage due to the fact they are most like real faces and are most likely to recruit holistic processing. A possible explanation for why I didn't find this interaction is simply because both Nord and Altmer faces are similarly different from real faces. Another possibility is that the performance benefit for Altmer faces outweighs any significant upright advantage a participant may have for Nord faces.

Three-way interaction. Expertise, race, and orientation significantly interacted for reaction time. The interaction suggests that experts showed a significant upright advantage for Nord faces over Altmer faces where novices did not. This aligns with my hypothesis that experts demonstrate trained expertise for faces they encounter most. However, there was no three-way interaction for accuracy. It is possible that this was skewed by the performance benefit for Altmer faces and that reaction time is a more sensitive measure for demonstrating expertise in this context. It is difficult to come to a strong conclusion without a corresponding interaction for accuracy, but this interaction provides evidence for the focus of Experiment 1; demonstrating that experts have generated expertise not only with Skyrim

faces generally but also with a specific, more commonly encountered subset of Skyrim faces.

Results for 1B. A measure was computed for each participant that represented the proportion of face-directed fixations in a video. Face proportion scores ranged from .01 (low number of face-fixations) to .35 (high number of face-fixations). I hypothesized that face-fixation measure would positively correlate with overall performance, upright advantage for performance, performance on Nord faces, and face-fixation measure would negatively correlate with overall reaction time and upright advantage for reaction time. I hypothesized that face-fixation measure would positively correlate with self-report of interaction frequency, self-reported importance of character recognition, and self-report of interaction frequency with Skyrim characters (for expert participants).

Face-fixation proportion was not significantly correlated with any behavioral measure, including overall performance, $r(93) = 0.105$, $p = .314$, overall reaction time, $r(93) = -0.11$, $p = .289$, upright advantage for performance, $r(93) = 0.16$, $p = .122$, upright advantage for reaction time, $r(93) = -0.08$, $p = .802$, or performance on Nord faces, $r(93) = 0.105$, $p = .263$. Face-fixation proportion was not correlated with other survey measures: including self-report of interaction frequency, $r(93) = -0.05$, $p = .610$, importance of character recognition, $r(93) = -0.11$, $p = .29$, or self-report of interaction frequency with Skyrim characters (for expert participants), $r(44)$

= 0.17, p = .268. I also compared face-fixation with self-reported looking strategies by dividing participants into two groups based on self-report of whether they looked at faces and comparing face-fixation proportions. The difference in face-fixation proportions between the two groups was not significant, $t(46)$ = -1.168, p = 0.125.

Discussion for 1B. There were no significant correlations between face-fixation proportion and behavioral or survey measures. This could suggest that there is no relationship between how often people look at faces and recognition advantage and that participants have inaccurate metacognitive judgements of how often they look at faces. Another possible explanation for why I was unable to reject any null hypothesis is that the video was not representative of what players report doing when they play. The benefit of eye-tracking with a video of gameplay over actual gameplay is that it allowed for a more streamlined analysis to be done. However, it is likely that participants view videos of gameplay differently than they view their actual gameplay. In future studies, recording eye-tracking when participants play Skyrim may be more informative than recording eye-tracking for a video of gameplay.

Results for 1C. Survey data were linked to the average accuracy for each participant. Answers to the open-ended questions were coded based on commonly-reported factors. An accuracy score was calculated for the Skyrim identity recognition task.

Video game habits. 62.7% of experts self-identified as gamers, compared to 5.56% of novices. 23% of experts used to self-identify as gamers but no longer do, compared to 14.81% of novices. Most novices (87.04%) reported spending 0-2 hours a week playing video games, whereas experts were fairly evenly spread across all options (see Table 3).

Table 3. How many hours a week do you spend playing video games?

	0-2	3-7	8-15	16-25	26+
Novices	87.04%	3.70%	9.26%	0	0.00%
Experts	15.69%	29.41%	23.53%	19.61%	11.76%

Experts were more likely to report starting to play video games at an earlier age (see Table 4), but across all participants, the reported starting age for playing video games did not significantly account for score.

Table 4. At what age did you start playing video games?

	1-4	5-9	10-13	14-17
Novices	7.41%	42.59%	40.74%	9.26%
Experts	19.61%	54.90%	23.53%	1.96%

Participants were asked to check their preferred type of interaction when playing video games with characters (see Table 5). Experts typically preferred Dialogue (84.31%), Quest and Task assistance (82.35%), and Destructive fighting (78.43%). Novices typically preferred Quest and Task assistance (59.26%) and Friendly competition (40.74%).

Table 5. What type of interactions do you prefer with the characters in the games you play?

	Friendly Dialogue	Bonding Competition	Quest/task Activities	Assistance	Tutorial	Romantic Interest	Destructive Fighting	Don't like interaction
Novices	38.89%	40.74%	14.81%	59.26%	27.78%	14.81%	27.78%	14.81%
Experts	84.31%	54.90%	43.14%	82.35%	43.14%	31.37%	78.43%	1.96%

Note. Participants could choose more than one answer

Independent sample t-tests based on $p < .05$ were computed for every type of interaction by dividing participants by whether they reported a variable as a preferred type of interaction and comparing accuracy. The only significant difference in average accuracy was for the Friendly Competition variable; participants who reported they preferred this interaction scored .039 higher than those who didn't, $t(103) = 2.098$, $p = .038$.

Participants were asked how important it was to recognize the characters in the games they played on a scale of 1 to 5; the average score was 3.56 with a Standard Deviation of 1.14. Self-report of importance of character recognition was not correlated with the average performance score.

Participants were asked how often they chose to interact with the characters in the games they play on a scale of 1 to 5; the average score was 3.4 with a Standard Deviation of 1.3. Self-report of interaction frequency was moderately positively correlated with self-report of importance of character recognition, $r(103) = .592$, $p < .0001$. However, self-report of interaction frequency was only marginally correlated with average performance score, $r(104) = 0.178$, $p = 0.069$ and not correlated with an upright advantage for performance, $r(104) = -0.13$, $p = 0.199$.

Skyrim habits. The average number of hours spent playing Skyrim for experts was 302.43 (n = 51). 62.75% of experts reported playing Skyrim in the last year. A one-way ANOVA compared performance and reaction times between groups of expert participants divided by when they last reported playing Skyrim (In the last 5 years, In the last year, In the last month, In the last week). There was no significant difference between groups for performance, $F(3,47) = 0.583$, $p = 0.629$, or reaction time $F(3,47) = 0.223$, $p = 0.879$. Among experts, performance on the face recognition task was moderately positively correlated with number of self-reported hours $r(51) = .3457$, $p < 0.05$ (see Figure 9).

Performance by number of hours for Experts, N=51

r=0.346p=0.013
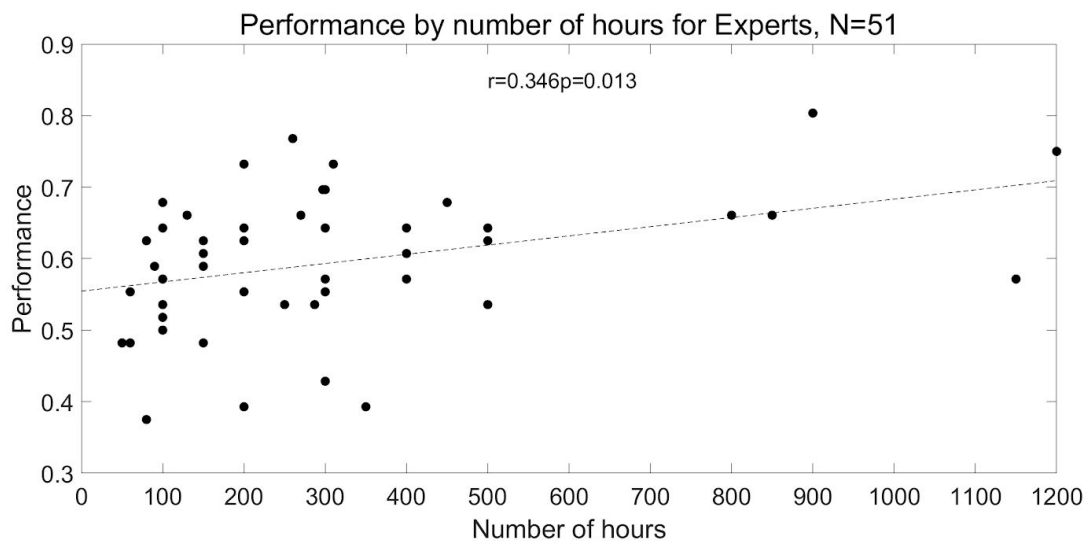
Figure 9. Correlation between number of self-reported hours and performance.

Experts were also asked to name their favorite character; 91.67% of experts reported having a favorite character (14.58% reported Lydia, 8.33%

reported Paarthurnax, 6.25% reported Sheogorath, 6.25% reported themselves).

Experts were asked to report how often they interacted with NPC's when playing Skyrim on a scale of 1 to 5; the average score was 3.71 with a Standard Deviation of 1.08. Self-report of interaction frequency in Skyrim did not significantly account for the variation in average performance score. A one way ANOVA showed that self-report of interaction frequency in Skyrim did account for upright advantage score, $F(4, 46) = 2.97$, $p = 0.029$. Post hoc comparisons using the Tukey HSD test indicated that the mean upright advantage score for the lowest level (1) of self-reported interaction (M = 0.38, SD = 0.15) was significantly higher than the mean upright advantage score for the highest level (5) of self-reported interaction (M = 0.089, SD = 0.14). There were no other significant differences between other groups.

Answers to the open-ended question "Why do you enjoy Skyrim?" were coded based on the following commonly-reported factors: Character Creation, Open world, Role-playing, Exploration, Freedom, Fantasy, Familiarity, Worldbuilding, Lore, and Immersion. Experts typically reported Freedom (n=20), Open world (n=17), and Exploration (n=12) as the reasons they enjoyed Skyrim.

Answers to the open-ended question "What is your favorite thing to do when playing Skyrim?" were coded based on the following commonly-reported factors: Exploration, Character building,

Creation/Construction, Quests, Collecting items, Combat, Game progression, Game Breaking, Character interaction, and Stealth. Experts typically reported Exploration (n=22) and Quests (n=20) as their favorite thing to do when playing Skyrim.

Answers to the open-ended question "What's the strangest thing you ever did in Skyrim?" were coded based on the following commonly-reported factors: Expectation/Reality Mismatch, Game breaking, Comical Gameplay, Macabre, Glitch, and Game-master. Experts typically reported Comical Gameplay (n=22), Macabre (n=12), and Expectation/Reality Mismatches (n=10) as the strangest thing they ever did in Skyrim.

Answers to the open-ended question "What is your favorite thing to do when playing Skyrim?" were coded based on the following commonly-reported factors: Fantasy, Combat, Collection/Acquisition, Game progression, Exploration, and Bonding. Experts typically reported Fantasy (n=12), Exploration (n=12), Combat (n=11), and Game progression (n=10) as their favorite things to do when playing skyrim.

Looking strategies. Participants were asked to report where they looked when they were communicating with a character in a video game. I hypothesized that participants who reported looking at the face would have significantly higher scores than those who didn't, however, the difference in scores was non-significant $t(90) = -0.938$, $p = 0.35$.

Character recognition task. Expert participants were asked to complete a character recognition task for some commonly encountered characters in Skyrim. Accuracy on the face recognition task was coded as how many Skyrim identities a participant was able to correctly name (out of a total of 10) to represent a performance score for expert participants. I hypothesized that character recognition score would positively correlate with overall performance and reaction time, upright advantage for performance and reaction time, performance on Nord faces, self-report of interaction frequency, self-reported importance of character recognition, and self-report of interaction frequency with Skyrim characters (for expert participants).

Character recognition score was not correlated with overall performance, $r(50) = .273$, $p = .055$, or reaction time $r(50) = .096$, $p = .508$. Character recognition score was not correlated with upright advantage for performance, $r(50) = -.144$, $p = .319$. Character recognition score was moderately positively correlated with performance on Nord faces, $r(50) = 0.312$, $p = .028$. Character recognition score was moderately positively correlated with upright advantage for reaction time, $r(50) = 0.30$, $p = .034$. Character recognition score was not correlated with self-report of interaction frequency, $r(48) = .175$, $p = .225$, self-reported importance of character recognition, $r(48) = .051$, $p = .725$, or self-report of interaction frequency with Skyrim characters (for expert participants), $r(48) = .221$, $p = .124$.

Discussion for 1C. Novices and experts differ in the following ways; proportion that self-identified as gamers, hours spent playing video games per week, and starting age of playing video games. However, none of these factors alone significantly account for a difference in performance score. This is important to address because there is a possibility that any of these could be a potential way for a participant to gain expertise.

Participants that self-identify as gamers, who report spending more hours playing video games, or who started playing video games at an earlier age could see an overall benefit of performance for digital faces due to increased experience with video game faces. However, the non-significant difference in performance indicates that general expertise does not seem to account for performance with Skyrim faces. In other words, the performance benefit with Skyrim faces seems to be specific to expertise with Skyrim.

Experts were asked to report the approximate number of hours they have spent playing Skyrim. Number of self-reported hours was positively correlated with accuracy, which provides evidence for the hypothesis that people generate expertise with video game expertise over time. Despite differences in gameplay, it is likely that the more time a player spends playing Skyrim the more expertise they will generate with the characters.

There was a non-significant difference in performance and reaction time when dividing experts into groups based on when they last played Skyrim. This suggests that the expertise developed by participants is

long-term; the expertise gained from gameplay is maintained even through years of not playing.

I asked participants to report their preferred type of interaction to get insight into how differences in type of interaction may account for expertise; for example, participants who report a preference for dialogue or bonding activities may see a benefit in performance due to increased meaningful experience with video game faces. The only significant finding was participants who reported preferring Friendly Competition scored higher than those who didn't. Friendly Competition had the most similar proportions for experts (54.90%) and novices (40.74%), which could indicate that Friendly Competition is the kind of interaction that people find most meaningful in gaming contexts (see Table 5).

Expert participants were also asked to answer questions about their experiences in Skyrim. When asked to report why they enjoyed Skyrim, many participants reported answers that tied into the idea of exploring an open world, and having the freedom to do whatever they wanted in the open world. This ties in closely into what participants typically reported as their favorite thing to do and their favorite experience in Skyrim.

Participants typically reported their favorite experiences when playing Skyrim as being able to go into another world (the game world) and have a life their with freedom to do whatever they wished. Participants typically reported enjoying the freedom to do things in game that they

couldn't in real life; they enjoyed being able to be whoever they wanted and doing whatever they wanted. Participants self-reported favorite experiences tied closely into the kinds of activities they enjoyed doing in Skyrim. They named what specific things they enjoyed doing when given the freedom to do so, such as discovering actions and consequences of actions. Participants reported liking discovering and interacting with aspects of Skyrim, and reported enjoying exploring and questing as their favorite activities. Answers to the open-ended questions tended to align with the gamer motivation model (Yee, 2006). Typically players reported enjoying Skyrim for the factors Achievement and Immersion, with a heavy emphasis on subcomponents of Advancement, Competition, Discovery, Role-playing, Customization, and Escapism (Yee, 2006).

I also asked expert participants to report the strangest thing they did in Skyrim. Typically participants reported embracing the funny moments in typical gameplay moments and when there was a mismatch between expectation and reality. The experiences participants reported as strange usually were tied into the idea of control, either by a participant losing control over something in the game, or feeling in control and creating strange and funny experiences for themselves. Participants also reported enjoying testing the limits of the game, either through game mechanics or glitches.

Expert participants were also asked to complete a character recognition task for some commonly encountered characters in Skyrim. The

score on this recognition task was positively correlated with performance on Nord faces, which could imply that increased knowledge of named characters helps increase expertise for Nord faces because many named characters are Nord (and very few are Altmer). The score on this recognition task was also positively correlated with upright advantage for reaction time, which implies an overall benefit for reaction time for participants who scored lower on the recognition task.

Discussion

Participants who had expertise with Skyrim scored significantly higher on accuracy on Skyrim faces than novice participants, which suggests an effect of expertise with Skyrim faces. Number of hours was positively correlated with this performance measure. These results suggest that video game players could gain expertise with specific video game faces (such as Skyrim) from playing hundreds of hours of a video game.

Participants who had expertise with Skyrim also showed an upright advantage in speed for a specific and commonly encountered race of Skyrim faces (Nord), where there was not one for a less commonly encountered race of Skyrim faces (Altmer). Novice participants did not show this difference, which suggests that participants who play Skyrim have generated expertise not only with Skyrim faces generally, but also with a specific, more commonly encountered subset of Skyrim faces.

The benefit of performance on Skyrim faces was only found for Skyrim experts; novice participants who reported playing other video games did not show a performance benefit. This suggests that there is no general expertise with video game faces; benefits in recognition for video game faces are specific to the particular style that people have experience with.

These results suggest that participants with no specific experience with Skyrim faces process the faces similar to other-race faces, whereas participants with specific experience with Skyrim faces process the faces similar to own-race faces. Gameplay seems to mediate this boundary by creating expertise with subsets of digital faces.

This experiment provides evidence that the recognition boundary between real and digital faces can be mediated by expertise, but is there also an effect of realism? Skyrim faces are stylized but still maintain what can be considered as typical human features (for Nord identities). An important next step is to investigate if realism also mediates the boundary between real and digital faces in novice participants. To explore this, Experiment 2 tested the difference in recognition scores between human faces, medium-realism digital representations of faces (as in Experiment 1), and high-realism digital representations of faces.

Chapter V

Experiment 2: Testing the Expertise Across Different Styles of

Digital Faces

Experiment 2 used an upright and inverted face recognition task to

test if participants with no video game experience showed increased

accuracy for faces that are more realistic over faces that are more stylized.

Morphed photographs of realistic faces should show the greatest upright

advantage (based on prior literature on the inversion effect).

Medium-realism digital representations of faces should show a robust

upright advantage similar to the one novices demonstrated in 1A; an upright

advantage of 0.176 in accuracy, and a speed advantage of .389 seconds. I

hypothesize that correlations of the scores for each participant between the

morph faces and each style will be modulated by how realistic the digital

representations are. High-realism digital representations of faces from the

video game Monster Hunter: World should show the highest correlation with

morph faces and low-realism digital representations of faces from the video

game Blade & Soul should show the lowest correlation with morph faces.

Method

Participants. Participants (n=58) were recruited through the SONA

system at University of California, Santa Cruz, and were given class credit in

exchange for their time. 5 participants were excluded from the analysis due

to low performance on Morph faces. For Experiment 2, only video game

65

novices were recruited. Video game novices were defined as participants with less than 50 hours of gameplay in The Elder Scrolls V: Skyrim, Monster Hunter: World, and Blade & Soul. 7 participants were excluded due to self-report of more than 50 hours of gameplay with one or more of the games. This left 46 participants for the main analysis.

Procedure. Participants came to the lab and completed a face recognition task (2A) and a survey (2B) on laboratory computers. Each participant used only their initials and date to identify themselves on the face recognition task and the survey.

Procedure for 2A. The behavioral task for Experiment 2 consists of an upright and inverted face recognition task, similar in design to Experiment 1. Faces were presented in blocks based on type of face: photos of morphed faces, high-realism digital representations of faces from the video game Monster Hunter World, medium-realism digital representations of faces from the video game The Elder Scrolls V: Skyrim, and low-realism digital representations of faces from the video game Blade & Soul (see Figure 4 for examples). A target digital face was displayed to the participants at one of two angle conditions (either front view or ¾ view), followed by a mask, followed by the target face and two distractor faces at the other angle. At random, half of the trials used upright faces, and half used inverted faces. Participants were asked to select the target face by pressing a corresponding key and reaction time and accuracy will be recorded.

Stimuli for 2A. The stimuli were morphed photos of faces created in Abrasoft FantaMorph 5 and avatar faces created in three video games (Monster Hunter: World, The Elder Scrolls V: Skyrim, and Blade & Soul).

The morph faces were constructed with Abrasoft FantaMorph 5 using Caucasian neutral male faces in front and ¾ view from the Radboud faces database (Langner, et al., 2010). For each identity, 3 variants were constructed by morphing the original identity with 3 new identities to create 3 distractor images (see *Figure 10A* for an example).

Stimuli for the digital representations were composed of 10 different preset identities and 3 variants. Screenshots of each of the 3 identity variants were taken at two views. In each case, two of the variants were used as distractors, and one was used as a target. The face stimuli was constructed with the default character creator in each game. Each identity was based on a different male preset (preset identities are constructed by game developers to provide additional starting points for character creation) available in each game's character creator.

The stimuli for the high-realism digital representations were avatar faces created in the role-playing video game Monster Hunter: World. For each preset in Monster Hunter: World wrinkles, facial hair, and makeup were removed. Three versions were constructed by changing the following preset face features; face, eyebrows, nose, and mouth. Screenshots of each variant

were taken with the face turned towards the front and ¾ to the right for a total of 6 images for each identity (see *Figure 10B* for an example).

The stimuli for the medium-realism digital representations were avatar faces created in the role-playing video game The Elder Scrolls V: Skyrim. The stimuli used are the same that were used in Experiment 1. For specific stimuli construction procedure, see section "Stimuli for 1A" in chapter IV (see *Figure 10C* for an example).

The stimuli for the low-realism digital representations were avatar faces created in the role-playing video game Blade & Soul. The chosen character race for the face stimuli was Lyn, as it was the most stylized option. For each preset in Blade & Soul, 3 versions were constructed. Variant 1 moved the slider to position 0 for each of the following features: brow position, brow angle, eye position, eye size, eye width, eye distance, nose position, nostril width, mouth position, mouth width, chin height, and chin width. Variant 2 moved the slider to a negative position between 4 and 12, and variant 3 moved the slider moved the slider to a positive position between 4 and 12. Screenshots of each variant were taken with the face turned towards the front and ¾ to the right for a total of 6 images for each identity (see *Figure 10D* for an example).
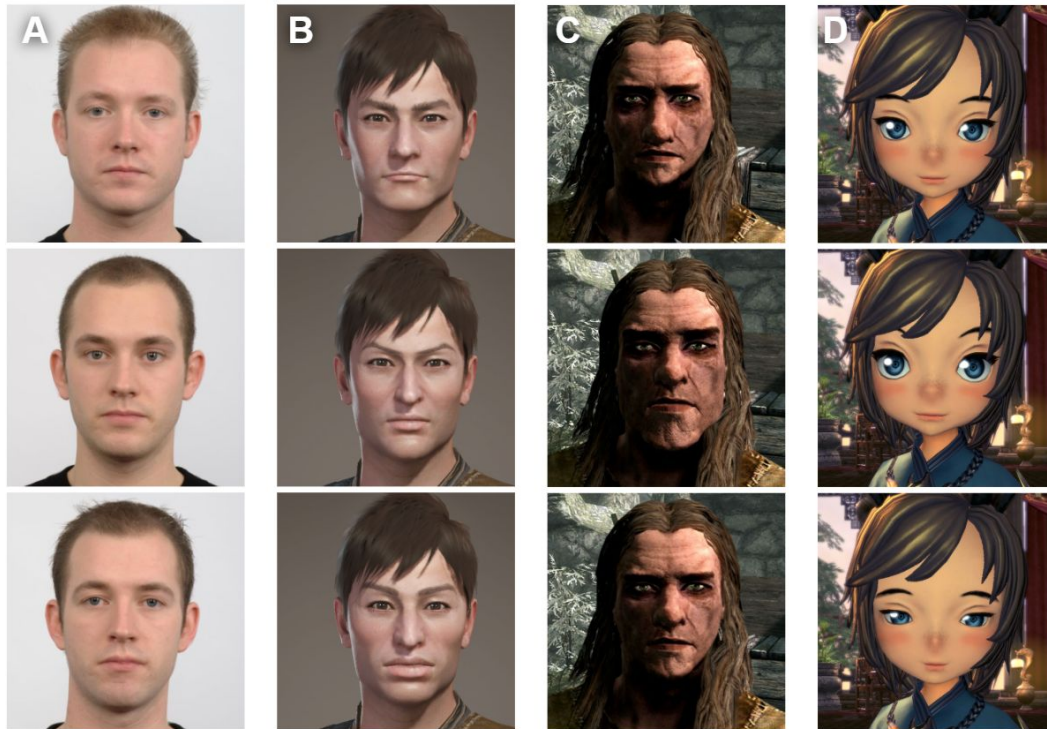
Figure 10: Examples of faces from each condition in Experiment 2. (A) Morph faces, (B) high-realism digital representations of faces from the video game Monster Hunter: World, (C) medium-realism digital representations of faces from the video game The Elder Scrolls V: Skyrim, (D) low-realism digital representations of faces from the video game Blade & Soul.

To account for possible differences in variability between avatar sets, a pairwise image variance analysis was conducted using the SSIM (Structural Similarity Index) function in Matlab (Wang, Bovik, Sheikh, & Simoncelli, 2004). The average SSIM score for The Elder Scrolls V: Skyrim faces is 0.72, the average SSIM score for the Blade and Soul faces is 0.76, the average SSIM score for the Monster Hunter World faces is 0.88, and the average SSIM score for the morph faces is 0.91. A one-way ANOVA showed significant differences

in mean scores, $F(3, 36) = 46.27$, $p < .001$. The analysis for Experiment 2 will account for this by performing a correlational analysis that compares the correlations between scores on real faces and scores on each style.

Procedure for 2B. 2B was a survey run with Google Forms after Part 2A of the experiment was completed. Participants were asked questions on the following topics: number of hours spent playing video games per week, age of first playing video games, self-identification of gamer stereotype, character recognition habits, game familiarity, and looking strategies during the behavioral task. For full survey see Appendix B.

Hypotheses. I predicted a main effect of style; I expected Morph faces to have the highest accuracy and lowest reaction times. I expected performance to drop as a style becomes less realistic; Monster Hunter faces should show worse scores than Morph faces but better scores than Skyrim or Blade & Soul faces. Skyrim faces should be worse than Morph or Monster Hunter faces, but better than Blade & Soul faces. I expected Blade & Soul faces to have the lowest accuracy. I predict a main effect of orientation; faces that are presented upright should have higher accuracy than faces shown inverted. I also predict an interaction between style and orientation; upright advantage should decrease the less realistic a face becomes. I expected the inverse for reaction time.

I predict that there will be no significant difference in scores on Skyrim (Nord) faces between novice participants in Experiment 1 and

participants in Experiment 2 because the sample populations are similar and the stimuli and procedure are identical. I predict rating of task difficulty to negatively correlate with accuracy. I expected difficulty ratings to be scaled based on SSIM score; higher similarity measures will have increased difficulty ratings. With this measure, Morph faces (SSIM: 0.91) should be most difficult, followed by Monster Hunter (SSIM: 0.88), Blade and Soul (SSIM: 0.76), and easiest should be Skyrim faces (SSIM: 0.72). I expect average accuracy to be positively correlated with the following survey measures: reported starting age for playing video games, self-report of importance of character recognition, and self-report of interaction frequency. I also expect some survey measures to be positively correlated with each other, such as self-report of interaction frequency and self-report of importance of character recognition.

Results

Results for 2A. Performance and reaction time in the face recognition task was calculated for each style by averaging scores across the within-subjects factor orientation. Upright advantage scores were calculated by averaging the difference between accuracy for upright and inverted faces. The standard tested was $p < .05$ and the pairwise comparisons were adjusted for Least Significant Difference.

A two-way MANOVA with two repeated measures was conducted to measure effects of style (within-subjects: Morphs, Monster Hunter, Skyrim,

Blade & Soul), and orientation (within-subjects: upright, inverted), and possible interactions between them on performance and reaction time (see Table 5). There was a significant effect of style, Wilks' Lambda = .312, $F(6, 40)$ = 14.677, $p < 0.001$. There was a significant effect of orientation, Wilks' Lambda = .2, $F(2, 44)$ = 87.918, $p < 0.001$. There was also a significant interaction between style and orientation, Wilks' Lambda = .669, $F(6, 40)$ = 3.301, $p = 0.01$.

Table 6. Multivariate Effects

Variable(s)	Wilks' Lambda	F	df	Error df	Sig.
Style	0.312	14.677[1]	6	40	0
Orientation	0.2	87.918[1]	2	44	0
Style * Orientation	0.669	3.301[1]	6	40	0.01

Note. [1]Exact Statistic

Style. There was a significant effect of style on performance at the $p < .05$ level, $F(1, 45)$ = 54.657, $p < .0001$. The average performance score on the face recognition task was highest for Monster Hunter faces (M=0.668, SD=.018), second highest for Morph faces (M=0.649, SD=.0124, third highest for Blade & Soul faces (M=0.558, SD=.016), and lowest for Skyrim faces (M=0.544, SD=.017) (see Figure 11). A follow up analysis was computed with a pairwise comparison of accuracy (see Table 6).
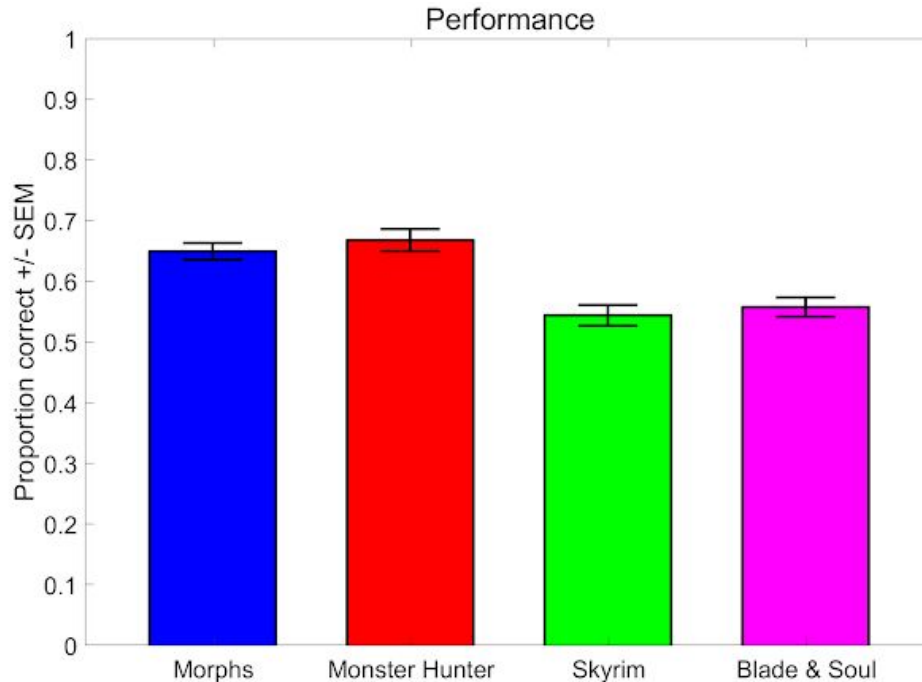
Figure 11. accuracy across style.

Table 7. *Pairwise Comparisons for Performance across Style*

	Mean Difference	Std. Error	Sig.[1]	95% Confidence Interval for Difference[1]	
				Lower Bound	Upper Bound
Morph - Monster Hunter	-0.018	0.017	1[1]	-0.064	0.027
Morph - Skyrim	.105*	0.017	0[1]	0.059	0.152
Morph - Blade & Soul	.092*	0.017	0[1]	0.044	0.14
Monster Hunter - Skyrim	.124*	0.017	0[1]	0.077	0.171
Monster Hunter - Blade & Soul	.110*	0.017	0[1]	0.062	0.159
Skyrim - Blade & Soul	-0.014	0.019	1[1]	-0.065	0.038

Note. Based on estimated marginal means

* The mean difference is significant at the .05 level.

[1] Adjustment for multiple comparisons: Bonferroni.

A pairwise comparison of accuracy between Morph and Skyrim faces indicated that average performance score for Morph faces (M=.649, SD=.014) was significantly higher than the average performance score for Skyrim faces (M=.544, SD=.017), $t(45)$ =6.21, p <.0001. A pairwise comparison of accuracy between Morph and Blade & Soul faces indicated that average performance

score for Morph faces (M=.649, SD=.014) was significantly higher than the average performance score for Blade & Soul faces (M=.558, SD=.016), $t(45)$ =5.28, p <.0001. A pairwise comparison of accuracy between Monster Hunter and Skyrim faces indicated that average performance score for Monster Hunter faces (M=.668, SD=.018) was significantly higher than the average performance score for Skyrim faces (M=.544, SD=.017), $t(45)$ =-7.26, p <.0001. A pairwise comparison of accuracy between Monster Hunter and Blade & Soul faces indicated that average performance score for Monster Hunter faces (M=.668, SD=.018) was significantly higher than the average performance score for Blade & Soul faces (M=.558, SD=.016), $t(45)$ =6.31, p <.0001. There was no significant difference between Morph and Monster Hunter faces, $t(45)$ =-1.11, p>.05. There was no significant difference between Skyrim and Blade & Soul faces, $t(45)$ =, p >.05.

I also computed a independent samples t-test between performance on Skyrim faces by participants in Experiment 2 and performance on Nord faces by participants in Experiment 1. There was a non-significant difference in performance between the two groups of participants, $t(93)$ =-1.77, p >.05 .

There was also a significant effect of style on reaction times at the p < .05 level, $F(1, 45) = 24.291$, $p < .0001$. The average reaction time on the face recognition task was fastest for Monster Hunter faces (M=2.08 seconds, SD=.069), second fastest for Morph faces (M=2.3 seconds, SD=.063), third fastest for Skyrim faces (M=2.43 seconds, SD=.1), and slowest for Blade &

Soul faces (M=2.48 seconds, SD=.079) (see Figure 12). A follow up analysis

was computed with a pairwise comparison of reaction times (see Table 7).
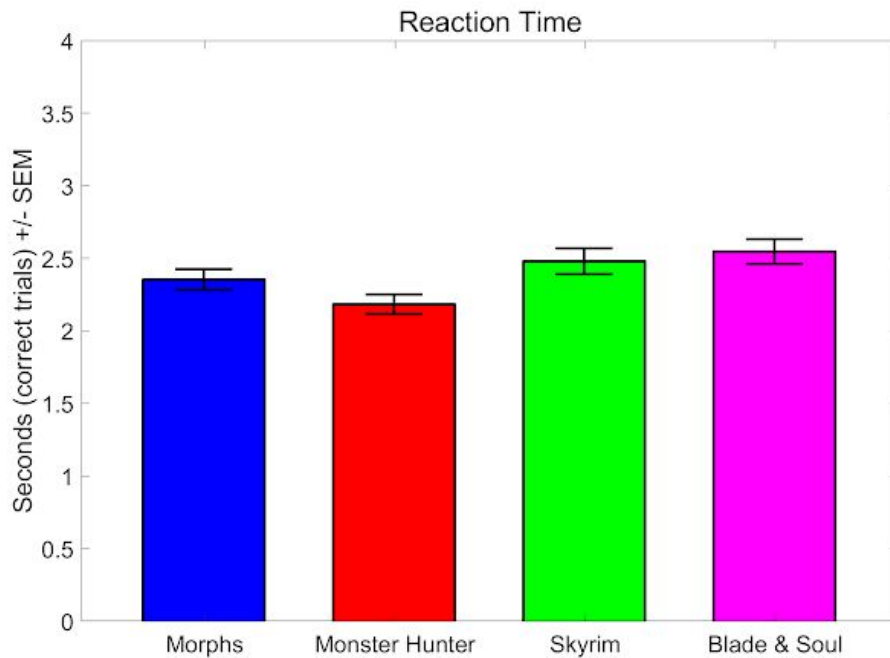
Figure 12. Reaction times across style.

Table 8. *Pairwise Comparisons for Reaction time across Style*

| | Mean Difference | Std. Error | Sig.[1] | 95% Confidence Interval for Difference[1] | |
				Lower Bound	Upper Bound
Morph - Monster Hunter	.220*	0.052	0.001[1]	0.076	0.364
Morph - Skyrim	-0.131	0.07	0.392[1]	-0.323	0.061
Morph - Blade & Soul	-.182*	0.062	0.03[1]	-0.352	-0.012
Monster Hunter - Skyrim	-.352*	0.084	0.001[1]	-0.583	-0.121
Monster Hunter - Blade & Soul	-.403*	0.064	0[1]	-0.578	-0.227
Skyrim - Blade & Soul	-0.051	0.094	1[1]	-0.31	0.208

Note. Based on estimated marginal means

* The mean difference is significant at the .05 level.

[1] Adjustment for multiple comparisons: Bonferroni.

A pairwise comparison of reaction times between Morph and Monster

Hunter faces indicated that average reaction time for Monster Hunter faces

(M=2.08 seconds, SD=.069) was significantly faster than the average reaction

time for Morph faces (M=2.3 seconds, SD=.063), $t(45)$ =3.49, p =.0011. A

pairwise comparison of reaction times between Morph and Skyrim faces

indicated that average reaction time for Morph faces (M=2.3 seconds,

SD=.063) was significantly faster than the average reaction time for Skyrim

faces (M=2.43 seconds, SD=.1), $t(45)$ =-2.39, p = 0.02. A pairwise comparison

of reaction times between Morph and Blade & Soul faces indicated that

average reaction time for Morph faces (M=2.3 seconds, SD=.063) was

significantly faster than the average reaction time for Blade & Soul faces

(M=2.48 seconds, SD=.079), $t(45)$ =-3.05, p = 0.003. A pairwise comparison

of reaction times between Monster Hunter and Skyrim faces indicated that

average reaction time for Monster Hunter faces (M=2.08 seconds, SD=.069)

was significantly faster than the average reaction time for Skyrim faces

(M=2.43 seconds, SD=.1), $t(45)$ =-7.26, p <.0001. A pairwise comparison of

reaction times between Monster Hunter and Blade & Soul faces indicated

that average reaction time for Monster Hunter faces (M=2.08 seconds,

SD=.069) was significantly faster than the average reaction time for Blade &

Soul faces (M=2.48 seconds, SD=.079), $t(45)$ = -5.74, p <.0001. There was no

significant difference in reaction time between Skyrim and Blade & Soul

faces, $t(45)$ = 0.947, p >.05. The reaction time results mirrored the

performance results.

I computed pairwise differences across style for upright faces and

inverted faces. For upright faces, accuracy for Morph faces were significantly

higher than Skyrim, $t(45) = 2.879$, $p < .01$, and Blade & Soul faces, $t(45) =$ 2.016172, $p > .05$, and not significantly different from Monster Hunter faces, $t(45) = .262$, $p > .05$. Upright accuracy for Monster Hunter faces were significantly higher than Skyrim, $t(45) = 2.752$, $p < .01$, but not significantly different from Blade & Soul faces, $t(45) = 1.638$, $p > .05$. There was no significant difference between Skyrim and Blade & Soul faces, $t(45) = -1.015$, $p > .05$.

For inverted faces, accuracy for Morph faces were significantly higher than Skyrim, $t(45) = 5.198$, $p < .001$, and Blade & Soul faces, $t(45) = 5.589$, $p < .001$, and not significantly different from Monster Hunter faces, $t(45) = -1.730$, $p > .05$. Upright accuracy for Monster Hunter faces were significantly higher than Skyrim, $t(45) = 6.995$, $p < .001$, and Blade & Soul faces, $t(45) = 7.567$, $p < .001$. There was no significant difference between Skyrim and Blade & Soul faces, $t(45) = -.045$, $p > .05$.

I also computed correlational analyses across style. Performance on Morph faces was positively correlated with performance on Monster Hunter faces, $r(46) = .495$, $p < .01$, Skyrim faces, $r(46) = .409$, $p < .01$, and Blade & Soul faces, $r(46) = .309$, $p < .05$ (see Table 8). However, an analysis that statistically compared the correlations revealed there was no significant difference between styles with how well they each correlated with performance on Morph faces. Reaction times for Morph faces were positively correlated with reaction time on Monster Hunter faces, $r(46) = .757$, $p < .01$, Skyrim faces,

$r(46)=.808$, $p<.01$, and Blade & Soul faces, $r(46)=.691$, $p<.01$ (see Table 9).

However, an analysis that statistically compared the correlations revealed

there was no significant difference between styles with how well they each

correlated with reaction times on Morph faces.

Table 9. Correlations of Performance Across Levels of Style

Style	1	2	3	4
1. Morph	-			
2. Monster Hunter	.495**	-		
3. Skyrim	.409**	.539**	-	
4. Blade & Soul	.309*	.486**	.355*	-

Note. ** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 10. Correlations of Reaction Time Across Levels of Style

Style	1	2	3	4
1. Morph	-			
2. Monster Hunter	.757**	-		
3. Skyrim	.808**	.680**	-	
4. Blade & Soul	.691**	.681**	.673**	-

Note. ** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Orientation. There was a significant effect of orientation on

performance at the $p < .05$ level, $F(1, 45) = 170.66$, $p < .001$. A pairwise

comparison of accuracy for the upright and inverted conditions indicated

that the average performance score for upright faces (M=.671, SD=.014) was

significantly higher than the average performance score for inverted faces

(M=.538, SD=.013), $t(45) =13.06$, $p <.0001$ (see Figure 13). On average,

participants scored .133 higher on accuracy for upright faces over inverted

faces; 95% confidence interval for difference [.112, .153].

There was a significant effect of orientation on reaction time at the $p <$

.05 level, $F(1, 45) = 16.21$, $p < .001$. A pairwise comparison of reaction time

scores for the upright and inverted conditions indicated that the average

reaction time for upright faces (M=2.23, SD=.059) was significantly lower

than the average reaction time for inverted faces (M=2.42, SD=.079), $t(104)$

=-3.13, $p <.01$ (see Figure 13). On average, participants responded .185

seconds faster for upright faces over inverted faces; 95% confidence interval

for difference [.092, .278].
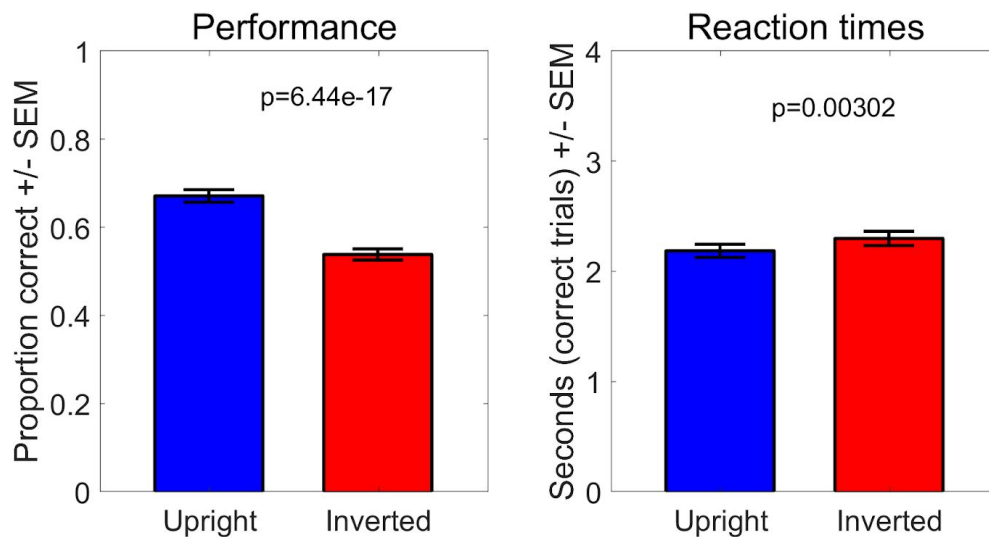
Figure 13: Average accuracy and reaction times for upright and

inverted faces.

Interactions. Interactions were calculated between style and

orientation for performance and reaction time. There was no significant

two-way interaction for reaction time: the interaction between style and orientation was non-significant, $F(1, 45) = .91$, $p > .05$. There was a significant interaction between style and orientation for performance at the $p < .05$ level, $F(1, 45) = 11.97$, $p < .01$ (see Figure 14). This interaction suggests that the difference upright and inverted faces that were more realistic (Morphs and Monster Hunter faces) showed a smaller inversion effect than those that are less realistic (Skyrim and Blade & Soul faces). The inversion effect for performance is biggest for Blade & Soul faces (M = 0.198, SD = 0.149), followed by Skyrim faces (M = 0.176, SD = 0.187), followed by Morph faces (M = 0.108, SD = 0.163), and smallest for Monster Hunter faces (M = 0.04, SD = 0.157).

A pairwise comparison of upright advantage accuracy between Morph and Blade & Soul faces indicated that average upright advantage for performance was significantly larger for Blade & Soul faces (M = 0.198, SD = 0.149) over Morph faces (M = 0.108, SD = 0.163), $t(45) = -2.728$, $p < .01$. A pairwise comparison of upright advantage accuracy between Monster Hunter and Skyrim faces indicated that average upright advantage for performance was significantly larger for Skyrim faces (M = 0.176, SD = 0.187) over Monster Hunter faces (M = 0.04, SD = 0.157), $t(45) = -3.591$, $p < .01$. A pairwise comparison of upright advantage accuracy between Monster Hunter and Blade & Soul faces indicated that average upright advantage for performance was significantly larger for Blade & Soul faces (M = 0.198, SD =

0.149) over Monster Hunter faces (M = 0.04, SD = 0.157), $t(45)$ =-4.65, p <.001. There was no significant difference in performance between Morph and Monster Hunter faces, $t(45) = 1.851943$, p >.05. There was no significant difference in performance between Morph and Skyrim faces, $t(45) =$ -1.745263, p >.05. There was no significant difference in performance between Skyrim and Blade & Soul faces, $t(45) = -0.628281$, p >.05.
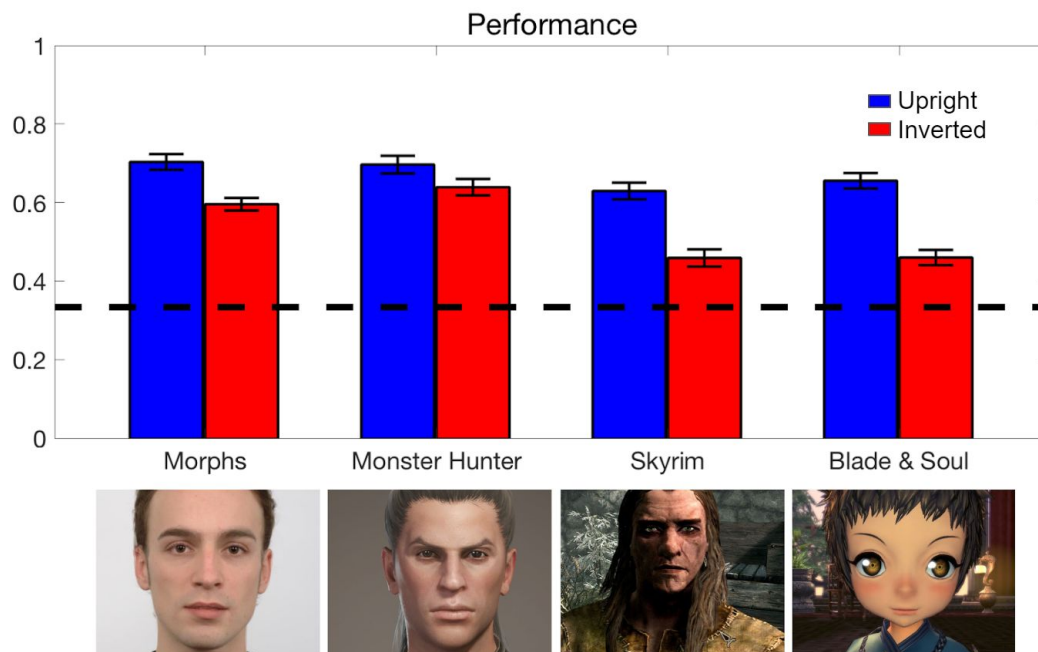
Figure 14: Average accuracy for upright and inverted faces, plotted by style of face.

Overall performance between upright and inverted faces is significantly positively correlated, $r(46) = 0.716$, $p < .001$. Overall reaction time between upright and inverted faces is significantly positively correlated, $r(46) = 0.836$, $p < .001$. Correlations between performance on upright and inverted faces were computed for each style of face. Upright performance

and inverted accuracy for Morph faces are not significantly correlated, $r(46)$ = 0.124, $p > .05$. Upright performance and inverted accuracy for Monster Hunter faces are significantly positively correlated, $r(46) = 0.421$, $p < .001$. Upright performance and inverted accuracy for Skyrim faces are not significantly correlated, $r(46) = 0.257$, $p > .05$. Upright performance and inverted accuracy for Blade & Soul faces are significantly positively correlated, $r(46) = 0.325$, $p < .05$.

Discussion for 2A. There were significant main effects for each independent variable and a significant interaction effect for both. In this section I discuss how these results relate to the hypotheses and address possible explanations.

Style. Participants tended to perform best on more realistic faces, performance was highest and reaction time was fastest for Morph and Monster Hunter faces, and the differences between the two were non-significant. Less realistic faces, such as the Skyrim and Blade & Soul faces, had lower accuracy and slower reaction times, but the differences between the two were non-significant. This initial result could provide evidence for a boundary between realistic and non-realistic digital representations.

I hypothesized that correlations between Morph faces and each style of digital face would be stronger the more realistic the digital faces became. Although the correlations seem to follow this trend initially, the differences

between the correlations are non-significant. It is possible that with more participants these differences would manifest more significantly.

The same stimuli were used for Nord faces in Experiment 1 and Skyrim faces from Experiment 2. The non-significant difference between performance on the Skyrim faces by novice participants in Experiment 1 and participants in Experiment 2 suggests that scores on this paradigm are robust.

Orientation. Participants performed better and were faster for upright faces over inverted faces. These results agree with the literature on face perception and provide evidence that the digital faces in this experiment are being processed similarly to faces. Performance on upright photos of faces is typically higher as a hallmark of gaining expertise only with faces that are oriented upright, and the inversion effect is often used in experimental paradigms to disrupt holistic processing. In the following sections, I discuss the differences in orientation effects across style.

Interactions. Differences in style were fairly consistent across orientation, but inverted faces tended to have larger mean differences between style and higher t scores. This suggests that much of the variability in performance across style comes from inverted performance, not upright. The interaction effect seems to be driven by the differences in performance between Morph faces and Blade & Soul faces, Monster Hunter faces and Skyrim faces, and Monster Hunter faces and Blade & Soul faces.

83

Faces that are more stylistic (i.e. Skyrim and Blade & Soul) tended to have larger upright advantages over faces that are more realistic (i.e. Morph and Monster Hunter) (see Figure 15). This is opposite of the prediction that more realistic faces should show a larger inversion effect because they are more likely to recruit face expertise from upright real faces, and therefore show a larger disadvantage for inverted faces. A possible explanation for these results is to do with looking strategies. It's possible that it is easier to pick up feature differences in more realistic representations, so participants were cued to look at more configural representations for more stylistic representations. Feature-based looking strategies would not see as great of a disadvantage from face inversion as configuration-based looking strategies because configuration-based looking strategies utilize more holistic processing and are more difficult to utilize for inverted faces. These differences are explored in the discussion section for 2B.

84
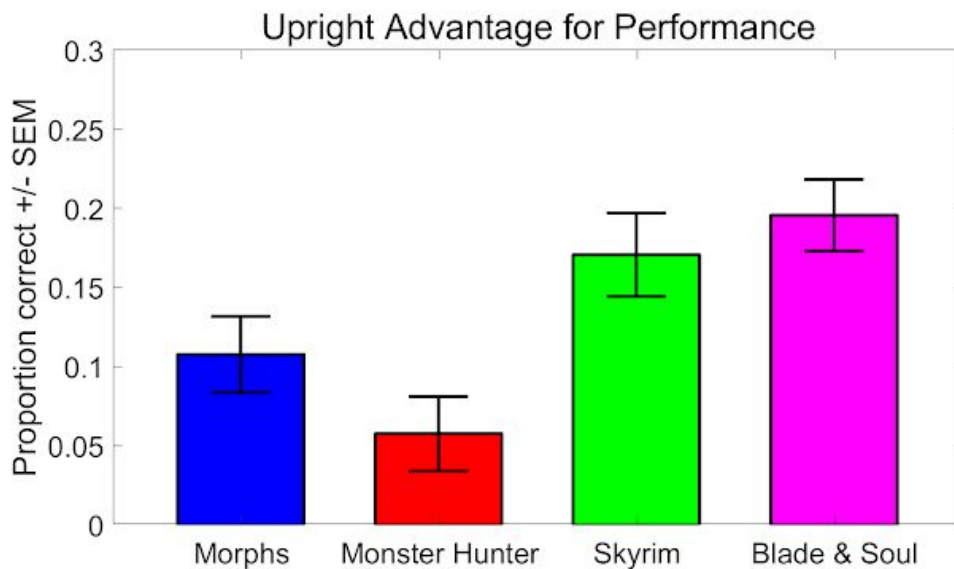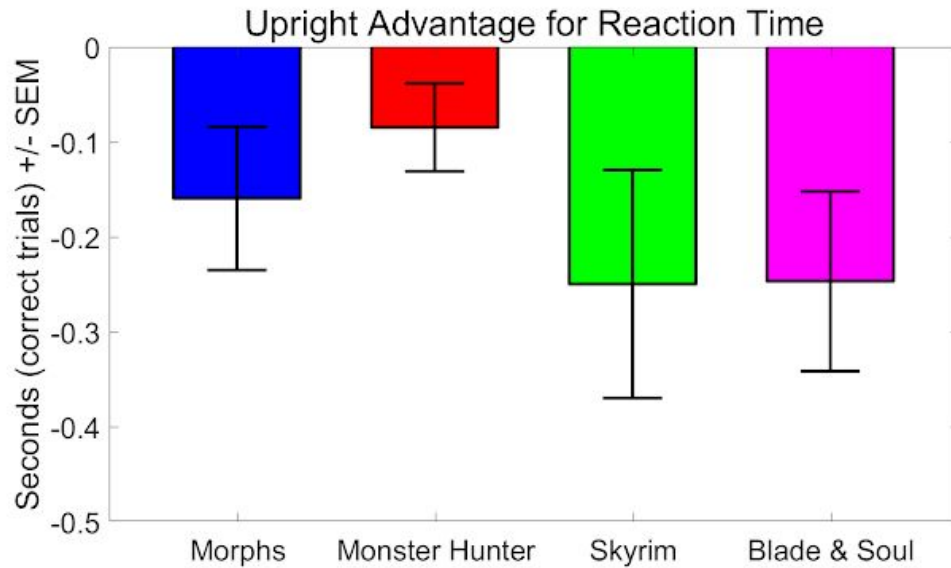
Figure 15. Upright advantage for performance shown as an increase in accuracy and decreased in reaction time across styles

The results from 2A suggest that there is a boundary between more realistic stylistic representations and less realistic representations, but do not show evidence that it occurs along a spectrum. The subtle differences between styles may be too difficult to pick up in order to define a spectrum of realism.

Results for 2B. Survey measures were grouped into two categories; video game habits and difficulty. Video game habits asked similar questions to those present in survey one, except without open-ended questions about play style. Difficulty ratings were scored from 1 to 5, with 5 coded as very difficult and 1 coded as very easy.

Video game habits. A one-way ANOVA between participants grouped by self-reported age of starting playing video games was not significant,

$F(4,41) =0.987$, $p > .05$. A one-way ANOVA between participants grouped by hours playing video games per week was not significant, $F(3,42) =2.206$, $p > .05$. A two-sample t-test between participants who reported looking at a face during communication and those who did not report looking at a face was non-significant, $t(44) = 1.248$, $p>.05$. Most participants reported looking at the text (n=46) or face (n=34) during communication with a character in a video game.

Overall accuracy were not significantly correlated with self-report of interaction frequency, $r(46) =.072$, $p >.05$, or self-report of character recognition importance, $r(46) =.102$, $p >.05$. However, interaction frequency and character recognition importance were moderately positively correlated $r(46) =.513$, $p <.001$. Participants typically reported their favorite interactions as Quest/task assistance (n = 42), Friendly competition (n = 37), and Destructive fighting (n = 33).

Difficulty. Difficulty ratings were analyzed as correlates for performance. Overall difficulty ratings did not significantly correlate with overall performance, $r(46) =-0.099$, $p >.05$. Ratings of difficulty for Morph faces did not significantly correlate with performance on Morph faces, $r(46) =.009$, $p >.05$. Ratings of difficulty for Monster Hunter faces did not significantly correlate with performance on Monster Hunter faces, $r(46) =.129$, $p >.05$. Ratings of difficulty for Blade and Soul faces did not significantly correlate with performance on Blade and Soul faces, $r(46) =.073$,

p >.05. However, difficulty ratings for Skyrim faces negatively correlated with performance on Skyrim faces, *r*(46) =-0.296, p < .05.

Morph faces (M= 3.65, SD = 1.099) were rated as significantly more difficult than Blade and Soul faces (M= 2.065, SD = 0.854), *t*(45) = 6.735, *p*<.001. Monster Hunter faces (M= 3.261, SD = 1.084) were rated as significantly more difficult than Blade and Soul faces (M= 2.065, SD = 0.854), *t*(45) = 7.325, *p*<.001, but significantly less difficult than Skyrim faces (M= 3.696, SD = 1.072), *t*(45) = -2.048, *p*<.05. Skyrim faces (M= 3.696, SD = 1.072) were rated as significantly more difficult than Blade and Soul faces (M= 2.065, SD = 0.854), *t*(45) = 9.367, *p*<.001.

Looking strategies. Participants were asked to report where they looked on a face for each style of face. For Morph faces, participants typically reported looking at the hair (n = 55). For Monster Hunter faces, participants typically reported looking at the mouth/lips (n = 42) and the eyebrows (n = 39). For Skyrim faces, participants typically reported looking at the chin (n = 45). For Blade & Soul faces, participants typically reported looking at the eye (n = 57).

Discussion for 2B. Participants were asked to report behavior pertaining to video game habits, perceived task difficulty, and looking strategies. The subfactors of self report of video game habits did not account for variation in overall performance score, but character recognition

importance did correlate positively with interaction frequency, similar to the result found in Experiment 1.

Aligning with the idea that higher SSIM scores will have increased difficulty ratings, difficulty ratings seemed to increase as stimuli became more realistic, with one exception. Skyrim faces were rated as most difficult, which could be due to a factor of the stimuli that isn't captured by the SSIM calculation. This factor could be face texture or shading. Skyrim is an older game than Monster Hunter or Blade & Soul; so the face texture is less smooth. The Skyrim character creator was also the only creator that didn't allow users to change ambient lighting; so the shadows on Skyrim faces are significantly more pronounced than the other face stimuli.

Looking strategies tended to differ between style. It is likely that participants are looking towards the most salient or distinctive feature for each style. The hair on the Morph faces is likely the most distinctive because it varied between face (unlike any of the digital faces, where the hair was kept consistent. Monster Hunter faces seemed to cue participants to look at the eyebrows and mouth, Skyrim faces seemed to cue participants to look at the chin, and Blade & Soul faces seemed to cue participants to look at the eyes. Each of these features aligns with what could be considered the most distinctive feature(s) in each style.

These differences could help explain the difference in inversion effect; the features that participants reported looking at for more realistic faces

88

(hair, eyebrows, and mouth) may suffer less of a disadvantage for inverted faces. The features participants reported looking at for more stylistic faces (chin and eyes) may tend to cause the configural information of the face to shift more, and these differences in configuration may explain the larger disadvantage for inverted faces. However, participants tended to report utilizing feature-based looking strategies overall; which could be a result of the experimental paradigm, the unfamiliarity with the faces, or both.

Discussion

These results suggest that realism of style, although less nuanced than hypothesized, does account for variation in score. More realistic styles tend to show increased performance and faster reaction times over less realistic styles of digital faces. This relates to the idea that more realistic representations of faces are better able to utilize the expertise we have for real faces. Less realistic renderings, however, may be less able to recruit real face expertise. This could be due to their relative position in face space; if we consider the dimension of face space to form due to expertise with real faces, highly stylistic representations are going to be represented farther away from the typical face. Practice discriminating these stylized faces may alter the dimensions of the space to include the new faces, and depending on the level of expertise, there may be dimensions specifically coded to that style.

Chapter VI

General Discussion

The results for experiments 1 and 2 provide evidence to suggest that expertise and style both mediate the boundary between processing for real and digital faces. As discussed in Chapter II, real and digital faces have a boundary between them that manifests as performance decreases for digital faces. The results of Experiment 1 and Experiment 2 suggest that the boundary is flexible, and differences in performance can be mediated by familiarity with digital faces and realism of digital faces.

It seems that video game players may be learning video game faces incidentally, as a result of narrative structure and interaction. This simulates a more natural approach to the development of expertise than traditional training studies and should be considered as a valid methodological approach to expertise research. In the context of these studies, expertise manifests as increased recognition scores and faster reaction times. In addition to expertise, it seems that more realistic styles are processed more similarly to real faces. This suggests that more realistic digital representations of faces may be able to recruit face processing mechanisms for real faces. More realistic faces may be relying on already developed face expertise, whereas less realistic face representations require additional experience to discriminate between identities.

Crookes et al. (2015) posited that there were three possible explanations for why people are worse at digital faces; (1) reduction in available information and relative similarity, (2) the lack of experience with digital faces, and (3) digital faces are automatically coded as outgroup faces. Experiment two provided tentative evidence against explanation 1; visual information was somewhat accounted for with the reduction of textural information in morph faces and the increase in textural information for Monster Hunter and Skyrim faces. SSIM scores also demonstrated that digital faces are actual less similar than morph faces. Experiment 1 provided evidence against explanation 3; Altmer faces (which should be least familiar to both novices and experts) showed the highest recognition score. The evidence from these experiments together best support explanation 1, that the reduced ability to process digital faces is due to relative experience.

These experiments help take the first step in defining factors that influence the differences between the processing of real faces and digital faces. This boundary seems to be flexible, and can move based on expertise and realism of style. In this discussion, I will discuss limitations, future directions, and applications.

Limitations. Experiment 1 was limited by the distinctiveness of the Altmer faces and the use of a gameplay video for eye-tracking. Altmer faces consistently had higher accuracy, despite my hypothesis that less experience with Altmer faces would manifest as lower scores. It is possible that Altmer

faces are constructed in a way that makes them more distinctive in the places people attend to during the study. Future studies could account for this in two ways (1) eye-tracking during the behavioral experiment, and (2) selecting a less commonly encountered race that doesn't visually differ as much. Eye-tracking during the behavioral experiment could allow investigation of where participants are looking to make judgements. Participants with expertise would likely process the face more holistically, whereas participants without expertise may focus on one salient feature (that may be exaggerated in Altmer faces). Another way to account for this limitation is to select a less commonly encountered race that doesn't visually differ as much, Altmers have a low rate of encounter but tend to be more exaggerated and angular than Nord faces. Some possible alternative races could be Orc (a fantasy race that is visually less angular), Redguard (a race that parallels African faces), or Breton (a half-elf race that is similar to Nord faces). Researchers could also construct a new race to fit their needs.

For the eye-tracking experiment I used a pre-recorded video of gameplay. A pre-recorded video of gameplay was ideal for eye-tracking as it allowed for control over the visual input (the visual scene was identical across participants), head stability (playing a video game could create more body motion), and ease of analysis. However, a possible reason I did not find any significant results for eye-tracking is that watching a video of gameplay is too different from actually playing the game. An important next step could

be comparing eye-tracking behavior for participants who primarily play or who primarily watch the game and compare differences in performance scores on face recognition tasks for participants who primarily play the game to those who primarily watch gameplay (such as Twitch viewers.).

Face-fixation proportion did not align with self-reported looking strategies, which is an indicator that what people report doing during gameplay and what they are doing which watching gameplay may mismatch. To account for this, researchers should eye-track participants as they play through an area of gameplay (possibly the tutorial). This will make automatic coding difficult due to the differences in gameplay style, but face-fixation scores will more accurate represent actual behavior when playing video games. In this scenario, gameplay style and interaction type could also be accounted for; for example researchers could track how participants are interacting with other characters instead of relying on self-report or preferred interaction type.

Experiment 2 was limited by inconsistent SSIM scores and difficulty ratings for each style of faces. Ratings of difficulty generally aligned with SSIM scores (except for Skyrim faces). This suggests that some faces may have been perceived as more difficult than others due to relative similarity, however the difficulty ratings did not account for variations in score (except for Skyrim faces). Style is a difficult factor to control for as every video game style is designed and implemented by a different team. For Skyrim faces the

relative similarity of faces accounted for self-report of difficulty, and difficulty negatively correlated with score. Skyrim faces seem to be an outlier style compared to the others; likely due to the game age. However, it also has one of the largest player bases with the most expertise. For this study it was important to strike a balance between expertise and style, but a future study could focus on just one of these factors.

Another factor to explore is realism ratings for the stimuli from participants, as judgement of realism of style for Experiment 2 was made by researchers and participant perception may differ. Some of the more realistic stimuli could also fall into the uncanny valley (Mori, MacDorman, & Kageki, 2012). The uncanny valley is defined as a eeriness of highly-realistic human representations that reduces likeability. Some highly realistic faces may be perceived as more creepy, but there has been no research to link the uncanny valley with a reduction in performance on identity recognition tasks.

In addition, the morph faces are still being digitally rendered, and there is evidence to show that morph faces are perceived as more typical, younger, and more attractive (Galton, 1878; Langlois & Roggman, 1990; Busey, 1998). A possible alternative to using morph faces is to adjust facial features in photos of faces (similar to how digital stimuli were constructed). Ideal distractors for face photo recognition tasks would be pictures of individuals who look very similar in real life (twins, siblings, or dopplegangers).

Future studies. The experiments in this project provide evidence for that the boundary for processing real and digital faces is mediated by expertise and style. Future studies could define this boundary even more, investigate specific factors that contribute to expertise (such as interaction type), and how expertise interacts with style.

Participants who have played many hours of a video game show an advantage for processing faces of that specific style. This suggests that people can generate expertise with digital faces if they spend over 50 hours playing a game. Future experiments should test this expertise effect with other video games to see if it replicates. An important factor to determine is how much the type of interaction contributes to generating expertise. For example, it may take a longer amount of playtime for a player to gain expertise with faces from a game with no named characters or narrative, or shorter amount of playtime for a player to gain expertise with faces from a game where they are interacting with real people's avatars. Yee, Bailenson, Urbanek, Chang, and Merget (2007) suggested that video games and digital spaces mimic social situations from real life. It is not too far-fetched to assume that the same social mechanisms that are driving face expertise in real faces may also drive expertise in digital faces.

Experiment 1 provided evidence that experts with Skyrim showed a benefit for that style, and participants who reported expertise with video games more generally did not show this effect. This suggests that expertise

with a specific video game manifests as specific expertise for that style of digital faces. However, more evidence to support this could best be generated through a study that tests across both expertise and style. Do Skyrim experts show benefits in performance for faces that were created for the same game franchise (for example, The Elder Scrolls IV: Oblivion)? Do Skyrim experts show benefits in performance for faces that were created within the same game engine (for example, Fallout 4)? Do Skyrim experts show benefits in performance for faces that are similar in realism (for example, L.A. Noire)?

Applications. Evidence from these experiments suggests that people generate expertise from specific styles, and that more realistic faces are able to better recruit processing for real faces. In this section, I will discuss potential applications for this research, in the form of using style to represent the self, creating representative video game populations, and how expertise with digital faces may transform the way people perceive real faces.

Faces are represented digitally in many different ways on a spectrum of realism. The way style is represented in digital faces can be thought of as occurring on along this spectrum, styles that are more realistic fall towards the right, and styles that are more iconic fall left (see Figure). Scott McCloud (1993) discussed style in terms of representation of the self; the less realistic a style is the more you can relate to it as a representation of yourself. He posited that cartoon imagery allows for universality in ways that more realistic styles don't. The results from Experiment 2 suggest that as realism

of the style decreases identity recognition performance decreases. This could

be because as the faces become more abstracted and iconic, they are less able

to recruit the expertise that is required to notice subtle differences between

faces. This also could be why more stylized representations of faces tend to

have more customization options. Not only is it fun to explore the

possibilities of character creation in a virtual space, it also allows users to

discriminate themselves in less subtle ways.
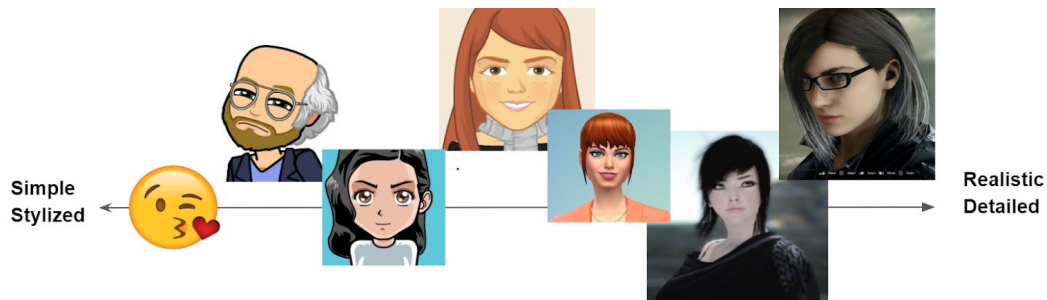
Figure 16. A theoretical spectrum of style for digital representations of faces.

This research can also be applicable for creating diverse and

representative video game character populations. As a style becomes less

realistic, it opens the possibility for more people to relate to the characters.

But if the races and populations aren't available during character creation or

represented in the world, it can feel alienating. Harrell and Harrell (2012)

discuss how well players feel an avatar from a video game represents them.

They highlight 3 axes that players typically refer to when constructing

characters; (1) stance towards avatar appearance, (2) stance towards avatar

ontological status, and (3) stance towards avatar use (Harrell & Harrell,

2012).

Avatar appearance occurs on a spectrum between the everyday (more typical or realistic features) and the extraordinary (more fantastical features). Avatar ontological status occurs on a spectrum between an accurate representation of real self and representation of an external character. Avatar use occurs on a spectrum between instrumental (created to perform a specific task) and identity play (using the avatar to embody a different self). It is important to have features that allow a player to construct a representation that aligns with their stance on each axis. Many character creators fall short of this, either the features needed to represent a wide audience are not available or they come with artifacts that are misaligned with character. For example, Redguards represent African faces, but choosing this race only allows the player to gain benefits for attributes related to combat abilities. It is important to have a range of races in video games that represent the population of players.

As technology becomes more integrated into people's daily lives, providing evidence for this becomes increasingly important. Digital representations of faces are a fundamental part of how we interact in digital spaces. Companies are adopting interfaces that utilize avatars and face representations and it is important to study how these interactions with digital faces could affect how we process faces "in the wild". This has interesting implications for younger populations who interact with these digital representations. It is possible that children are generating expertise

with digital faces that is manifesting not only as increased recognition ability with faces of that style, but also with other digital faces.

Developing expertise with digital representations of faces could create more flexible face representations. Video game players typically gain expertise with a wide variety of games and their different stylistic representations of faces. This increase in expertise with a large set of different digital face representations could increase the player's ability to learn and integrate new faces. Learning sets of new faces to understand game narrative could encourage individuation strategies that are independent of style (or race) coding. This practice, in turn, could make players less vulnerable to other-race effects in real life, and provides promising results for learning new races.

Appendix A

Survey for Experiment 1

ORE(1) post-experiment survey

Please answer the following questions.

* Required

1. **Subject initials** *

   _____

2. **Please enter today's date** *

   _____

   Example: December 15, 2012

3. **Did you participate in the eye-tracking study?** *

   Mark only one oval.

   ( ) Yes

   ( ) No

Video game habits

4. **How many hours a week do you spend playing video games?** *

   Mark only one oval.

   ( ) 0-2

   ( ) 3-7

   ( ) 8-15

   ( ) 16-25

   ( ) 26+

5. **At what age did you start playing video games?** *

   Mark only one oval.

   ( ) 1-4

   ( ) 5-9

   ( ) 10-13

   ( ) 14-17

   ( ) 18+

6. **Do you self-identify as a gamer?** *

   Mark only one oval.

   ( ) Yes

   ( ) No

   ( ) I used to, but not anymore

Character recognition habits

7. **How important is it to recognize the characters in the games you play?** *
   Mark only one oval.

	1	2	3	4	5	
Not important	◯	◯	◯	◯	◯	Very important

8. **How often do you choose to interact with the characters in the games you play?** *
   Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Very often

9. **What type of interactions do you prefer with the characters in the games you play? (You may select more than 1)** *
   Check all that apply.

   - ☐ Dialogue
   - ☐ Friendly competition
   - ☐ Destructive fighting
   - ☐ Bonding activities
   - ☐ Romantic interest
   - ☐ Quest/task assistance
   - ☐ Tutorial/knowledge acquisition
   - ☐ I do not like interacting with characters

10. **When you are communicating with a character in a video game, where are you looking?** *
    Check all that apply.

    - ☐ At the text
    - ☐ At the character's face
    - ☐ At the character's body
    - ☐ At the background
    - ☐ Other:

The Elder Scrolls V: Skyrim

11. **Have you played The Elder Scrolls V: Skyrim?** *
    Mark only one oval.

    - ◯ Yes
    - ◯ No   *Stop filling out this form.*

Skyrim habits

101

12. **How many hours have you played Skyrim?** *

_____

13. **When was the last time you played Skyrim?** *
   Mark only one oval.

   ◯ In the last week

   ◯ In the last month

   ◯ In the last year

   ◯ In the last 5 years

   ◯ More than 5 years ago

14. **What platform(s) do you play Skyrim on?** *
   Check all that apply.

   ☐ PC

   ☐ Xbox 360

   ☐ Xbox One

   ☐ PlayStation 3

   ☐ PlayStation 4

   ☐ Nintendo Switch

15. **How often do you interact with NPC's (non-playable characters) in Skyrim?** *
   Mark only one oval.

	1	2	3	4	5	
As little as possible	◯	◯	◯	◯	◯	As much as possible

Personal significance of narrative and gameplay

16. **Why do you enjoy Skyrim?** *

   _____

   _____

   _____

   _____

   _____

17. **What is your favorite thing to do when playing Skyrim?** *

   _____

   _____

   _____

   _____

   _____

18. **Who is your favorite Skyrim character and why?** *

_____

_____

_____

_____

_____

_____

19. **What was your favorite experience playing Skyrim?** *

_____

_____

_____

_____

_____

_____

20. **What's the strangest thing you ever did in Skyrim?** *

_____

_____

_____

_____

_____

Character recognition task

Please complete the character recognition task by following this link:
https://qtrial2018q3az1.az1.qualtrics.com/jfe/form/SV_8iYkm51eOhmYsTj

21. **Enter your completion code below.** *

_____

ORE(2) post-experiment survey

Please answer the following questions.

* Required

1. **Subject initials** *

   _____

2. **Please enter today's date** *

   _____

   Example: December 15, 2012

Video game habits

3. **How many hours a week do you spend playing video games?** *
   Mark only one oval.

   ◯ 0-2
   ◯ 3-7
   ◯ 8-15
   ◯ 16-25
   ◯ 26+

4. **At what age did you start playing video games?** *
   Mark only one oval.

   ◯ 1-4
   ◯ 5-9
   ◯ 10-13
   ◯ 14-17
   ◯ 18+

5. **Do you self-identify as a gamer?** *
   Mark only one oval.

   ◯ Yes
   ◯ No
   ◯ I used to, but not anymore

Character recognition habits

6. **How important is it to recognize the characters in the games you play?** *

Mark only one oval.

	1	2	3	4	5	
Not important	◯	◯	◯	◯	◯	Very important

7. **How often do you choose to interact with the characters in the games you play?** *

Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	Very often

8. **What type of interactions do you prefer with the characters in the games you play? (You may select more than 1)** *

Check all that apply.

- ☐ Dialogue
- ☐ Friendly competition
- ☐ Destructive fighting
- ☐ Bonding activities
- ☐ Romantic interest
- ☐ Quest/task assistance
- ☐ Tutorial/knowledge acquisition
- ☐ I do not like interacting with characters

9. **When you are communicating with a character in a video game, where are you looking?** *

Check all that apply.

- ☐ At the text
- ☐ At the character's face
- ☐ At the character's body
- ☐ At the background
- ☐ Other:

Game familiarity

10. **Have you played The Elder Scrolls V: Skyrim?** *

Mark only one oval.

- ◯ Yes, more than 50 hours
- ◯ Yes, less than 50 hours
- ◯ No

11. **Have you played Blade & Soul?** *

Mark only one oval.

- ( ) Yes, more than 50 hours
- ( ) Yes, less than 50 hours
- ( ) No

12. **Have you played Monster Hunter World?** *

Mark only one oval.

- ( ) Yes, more than 50 hours
- ( ) Yes, less than 50 hours
- ( ) No

Participation

13. **How difficult did you find the task overall?** *

Mark only one oval.

	1	2	3	4	5	
Very easy	( )	( )	( )	( )	( )	Very difficult

14. **How difficult was it to recognize faces from The Elder Scrolls V: Skyrim?** *

See example below

Mark only one oval.

	1	2	3	4	5	
Very easy	( )	( )	( )	( )	( )	Very difficult

106

15. **What attributes of faces from The Elder Scrolls V: Skyrim did you attend to most?** *

Check all that apply.

- [ ] Eyes
- [ ] Nose
- [ ] Mouth/lips
- [ ] Eyebrows
- [ ] Chin
- [ ] Hair
- [ ] Emotion/expression
- [ ] Other: _____

16. **How difficult was it to recognize faces from Blade & Soul?** *

See example below

Mark only one oval.

	1	2	3	4	5	
Very easy	◯	◯	◯	◯	◯	Very difficult

17. **What attributes of the faces from Blade & Soul did you attend to most?** *
Check all that apply.

- ☐ Eyes
- ☐ Nose
- ☐ Mouth/lips
- ☐ Eyebrows
- ☐ Chin
- ☐ Hair
- ☐ Emotion/expression
- ☐ Other: _____

18. **How difficult was it to recognize faces from Monster Hunter World?** *
See example below

Mark only one oval.

	1	2	3	4	5	
Very easy	◯	◯	◯	◯	◯	Very difficult

19. **What attributes of the faces from Monster Hunter World did you attend to most?** *

Check all that apply.

☐ Eyes

☐ Nose

☐ Mouth/lips

☐ Eyebrows

☐ Chin

☐ Hair

☐ Emotion/expression

☐ Other: _____

20. **How difficult was it to recognize morph faces?** *

See example below

Mark only one oval.

	1	2	3	4	5	
Very easy	◯	◯	◯	◯	◯	Very difficult

21. **What attributes of the morph faces did you attend to most?** *

Check all that apply.

☐ Eyes

☐ Nose

☐ Mouth/lips

☐ Eyebrows

☐ Chin

☐ Hair

☐ Emotion/expression

☐ Other: _____

References

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., & Lee, K. (2010).

Categorization, categorical perception, and asymmetry in infants'

representation of face race. *Developmental science*, *13*(4), 553-564.

Bailey, A. J., Braeutigam, S., Jousmäki, V., & Swithenby, S. J. (2005). Abnormal

activation of face processing systems at early and intermediate

latency in individuals with autism spectrum disorder: a

magnetoencephalographic study. *European Journal of*

Neuroscience, *21*(9), 2575-2585.

Banks, J., & Bowman, N. D. (2016). Avatars are (sometimes) people too:

Linguistic indicators of parasocial and social ties in player–avatar

relationships. *New Media & Society*, *18*(7), 1257-1276.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in

own-race face processing. *Psychological science*, *17*(2), 159-163.

Bartrip, J., Morton, J., & De Schonen, S. (2001). Responses to mother's face in

3-week to 5-month-old infants. *British Journal of Developmental*

Psychology, *19*(2), 219-232.

Bethesda Game Studios. (2011). The Elder Scrolls V: Skyrim [Video

game]. Rockville: Bethesda Softworks.

Blass, E. M., & Camp, C. A. (2004). The ontogeny of face identity: I. Eight-to

21-week-old infants use internal and external face features in

identity. *Cognition*, *92*(3), 305-327.

Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in

    the recognition of faces of own-and other-race persons. *Journal of*

    social issues, *41*(3), 139-155.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity:

    the power of an expertise framework. *Trends in cognitive sciences*,

    10(4), 159-166.

Busey, T. A. (1998). Physical and psychological representations of faces:

    Evidence from morphing. *Psychological Science, 9*(6), 476-483.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological

    evidence for configural processing in fingerprint experts. *Vision*

    research, *45*(4), 431-448.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants:

    Learning and memory. *Infant and Child Development: An*

    International Journal of Research and Practice, *10*(1-2), 67-74.

Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race

    faces: Implications for the representation of race in face space.

    Psychonomic Bulletin & Review, 11(4), 735-741.

Calder, A., Rhodes, G., Johnson, M., & Haxby, J. (Eds.). (2011). *Oxford*

    handbook of face perception. Oxford University Press.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis

    of the own-race bias in face recognition. *The Quarterly Journal of*

    Experimental Psychology Section A, *48*(4), 879-894.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No

 childhood development of holistic processing, novel face encoding, or

 face-space. *Cognition*, *111*(2), 219-247.

Crookes, K., Ewing, L., Gildenhuys, J. D., Kloth, N., Hayward, W. G., Oxner, M., ...

 & Rhodes, G. (2015). How well do computer-generated faces tap face

 expertise?. *PloS one, 10*(11), e0141353.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect

 of expertise. *Journal of Experimental Psychology: General*, *115*(2),

 107.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar

 and unfamiliar faces from internal and external features: Some

 implications for theories of face recognition. *Perception*, *8*(4),

 431-439.

Galton, F. (1879). Composite Portraits, Made by Combining Those of Many

 Different Persons Into a Single Resultant Figure. *The Journal of the

 Anthropological Institute of Great Britain and Ireland, 8,* 132-144.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise.

 Current opinion in neurobiology, *11*(2), 219-224.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for

 cars and birds recruits brain areas involved in face recognition.

 Nature neuroscience, *3*(2), 191.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring

mechanisms for face recognition. *Vision research*, *37*(12), 1673-1682.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object

recognition: bridging brain activity and behavior. *Journal of*

Experimental Psychology: Human Perception and Performance, *28*(2),

431.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., (1999).

Activation of the middle fusiform 'face area' increases with expertise

in recognizing novel objects. *Nat. Neuroscience 2*, 568–573.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble'

experts: a framework for studying expert object recognition

processes. *Vision research*, *38*(15-16), 2401-2428.

George, N., Driver, J., & Dolan, R. J. (2001). Seen gaze-direction modulates

fusiform activity and its coupling with other brain areas during face

processing. *Neuroimage*, *13*(6), 1102-1112.

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential

responses in the fusiform region to same-race and other-race faces.

Nature neuroscience, *4*(8), 845.

Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face

recognition: Reduction of the other-race effect. *Bulletin of the*

Psychonomic Society, *23*(3), 211-214.

Gomez, J., Barnett, M., & Grill-Spector, K. (2019). Novel childhood experience

suggests eccentricity drives organization of human visual cortex.

bioRxiv, 415729.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area

    subserves face perception, not generic within-category identification.

    Nature neuroscience, *7*(5), 555.

Guegan, J., Moliner, P., & Buisine, S. (2015). Why are online games so

    self-involving: A social identity analysis of massively multiplayer

    online role-playing games. *European Journal of Social Psychology*,

    45(3), 349-355.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the

    other-race effect in face recognition. *British Journal of Psychology*,

    99(1), 45-56.

Harrell, D. F., & Harrell, S. V. (2012). Imagination, computation, and

    self-expression: Situated character and avatar mediated identity.

    Leonardo electronic almanac, 17(2).

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human

    neural system for face perception. *Trends in cognitive sciences*, *4*(6),

    223-233.

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face

    recognition by shifting attention. *The Quarterly Journal of*

    Experimental Psychology, *59*(6), 996-1002.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and

    individuation in the cross-race recognition deficit: Toward a solution

to an insidious problem. *Journal of Experimental Social Psychology*,
43(2), 334-340.

Isbister, K. (2016). *How games move us: Emotion by design*. Mit Press.

Johnson, M. H. (2011). Face perception: A developmental perspective. *The
Oxford handbook of face perception*, 3-14.

Johnson, M. H., Grossmann, T., & Farroni, T. (2008). The social cognitive
neuroscience of infancy: Illuminating the early development of social
brain functions. In *Advances in child development and behavior* (Vol.
36, pp. 331-372). JAI.

Kang, S., Watt, J., & Isbister, K. (2006). The effect of static anthropomorphic
images on emotion perception in mobile-phone communication. In
Proc. 9th International Workshop on Presence.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a
module in human extrastriate cortex specialized for face perception.
Journal of neuroscience, *17*(11), 4302-4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region
specialized for the perception of faces. *Philosophical Transactions of
the Royal Society of London B: Biological Sciences*, *361*(1476),
2109-2128.

Kätsyri, J. (2018). Those virtual people all look the same to me:
Computer-rendered faces elicit a higher false alarm rate than
real human faces in a recognition memory task. *Frontiers in*

psychology, 9, 1362.

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality?. *Journal of experimental child psychology, 104*(1), 105-114.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science, 18*(12), 1084-1089.

Kleinhans, N. M., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., & Aylward, E. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain*, *131*(4), 1000-1012.

Krigolson, O. E., Pierce, L. J., Holroyd, C. B., & Tanaka, J. W. (2009). Learning To become an expert: Reinforcement learning and the acquisition Of perceptual expertise. *Journal of Cognitive Neuroscience*, *21*(9), 1833-1840.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological science, 1*(2), 115-121.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24(8)*, 1377—1388. DOI: 10.1080/02699930903485076

Lee, K., Anzures, G., Quinn, P. C., Pascalis, O., & Slater, A. (2011).

> Development of face processing expertise. In *Oxford handbook of
> face perception*.

Levin, D. T. (1996). Classifying faces by race: The structure of face

> categories. *Journal of Experimental Psychology: Learning, Memory,
> and Cognition*, *22*(6), 1364.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The

> ambiguous race face effect. Psychology, Public Policy, and Law,
> 7(1), 98.

Manning, P. (2009). Can the avatar speak?. *Journal of Linguistic

> Anthropology*, *19*(2), 310-325.

Nowak, K. L., Pandžić, I. S., Zorić, G., Šmid, K., Grundnig, S., Petri, T., &

Reinecke, L. (2008, January). The Psychology of Avatars: Real Life Effects

> of Virtual Communication. In *Proceedings of the 11th Annual
> International Workshop on Presence, Padua, Italy*.

Martey, R. M., & Consalvo, M. (2011). Performing the looking-glass self:

> Avatar appearance and group identity in Second Life. *Popular
> Communication*, *9*(3), 165-180.

Maurer, D., & Salapatek, P. (1976). Developmental changes in the scanning

> Of faces by young infants. *Child development*, 523-527.

McKone, E., Brewer, J. L., MacPherson, S., Rhodes, G., & Hayward, W. G.

> (2007). Familiar other-race faces show normal holistic processing

and are robust to perceptual stress. *Perception*, *36*(2), 224-248.

McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive neuropsychology*, *29*(1-2), 174-212.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces?. *Trends in cognitive sciences*, *11*(1), 8-15.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*(7), 608-615.

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanaka, J. W., & Zhou, G. (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception*, *39*(8), 1065-1085.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98-100.

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own-and cross-race face spaces. *Cognition, 116*(2), 283-288.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life?. *Science*, *296*(5571), 1321-1323.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, *18*(1), 79-85.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the national academy of sciences*, *102*(14), 5297-5300.

Ratan, R., & Hasler, B. S. (2010). Exploring self-presence in collaborative virtual teams. *PsychNology Journal*, *8*(1).

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological science*, *14*(4), 302-308.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioral tasks. *Cognition*, *103*(1), 34-79.

Rossion, B., & Curran, T. (2010). Visual expertise with pictures of cars correlates with RT magnitude of the car inversion effect. *Perception*, *39*(2), 173-183.

Scarborough, J. K., & Bailenson, J. N. (2014). Avatar Psychology. *Mark Grimshaw (Hg.): The Oxford Handbook of Virtuality, Oxford*, 129-144.

Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific reports, 7*, 45003.

Scott, L. S. (2011). Face perception and perceptual expertise in adult and

    developmental populations. *AJ, Calder, G., Rhodes, MH, Johnson, JV*

    Haxby,(Eds.), Oxford handbook of face perception, 195-213.

Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception.

    Psychological Science, *20*(6), 676-680.

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of

    the development of perceptual discrimination. *Current directions in*

    psychological science, *16*(4), 197-201.

Skyrim: Characters by Race. (n.d.). Retrieved from

    https://elderscrolls.fandom.com/

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face

    expertise. *Journal of Experimental Psychology: General*, *130*(3), 534.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object

    recognition. *Psychological science*, *12*(1), 43-47.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the

    basic level in the eye of the beholder?. *Cognitive psychology*, *23*(3),

    457-482.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008).

    Understanding evaluation of faces on social dimensions. *Trends in*

    cognitive sciences, 12(12), 455-460.

Tychsen, A., Hitchens, M., & Brolund, T. (2008). Character play: the use of

    game characters in multiplayer role-playing games across platforms.

Computers in Entertainment (CIE), *6*(2), 22.

Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *Neuroimage*, *15*(2), 454-461.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing, 13*(4), 600-612.

Williams, D., Kennedy, T. L., & Moore, R. J. (2011). Behind the avatar: The patterns, practices, and functions of role playing in MMOs. *Games and Culture*, *6*(2), 171-200.

Van Der Heide, B., Schumaker, E. M., Peterson, A. M., & Jones, E. B. (2013). The Proteus effect in dyadic communication: Examining the effect of avatar appearance in computer-mediated dyadic interaction. *Communication Research*, *40*(6), 838-860. Yee, N. (2006). Motivations for play in online games. *CyberPsychology & behavior, 9*(6), 772-775.

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior*, *10*(1), 115-121.

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special?. *Vision research*, *46*(22), 3802-3811.