THE PROCESSING OF CARTOONY AND PHOTOREALISTIC FACES

by

LIA KENDALL

B.Sc., The University of Toronto, 2013

M.A., The University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2019

© Lia Kendall 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

The processing of cartoony and photorealistic faces

submitted by Lia Kendall in partial fulfillment of the requirements for

the degree of Doctor of Philosophy in

Psychology

Examinig Committee:

Alan Kingstone, Psychology

Co-supervisor

Rebecca Todd, Psychology

Co-supervisor

Jess Tracy, Psychology

Supervisory Committee Member

Toni Schmader, Psychology

University Examiner

Molly Babel, Linguistics

University Examiner

Anthony Lambert

External Examiner
Abstract
Cartoony faces are everywhere, from texting apps to children’s cartoons. Cartoony faces are also often used in cognitive research contexts, where they are used to stand in as simplified photographic faces for their ease of manipulation and creation. This presents an often unspoken assumption: that a cartoony face is analogous to a photographic face in how it will be responded to, understood, and processed by the brain. Over 8 experiments, my dissertation aims to better understand how cartoony faces are similar to photographic faces, and where they differ. In the first two experiments, I found that there was no evidence that people see themselves in simple cartoony faces, as had been suggested in the past, and also that participants associated their photographs more with themselves than drawings of themselves. In Experiments 3 and 4, I found that, as faces become more cartoonized, they become easier to discriminate expressions on as well, and that such changes to ‘cartoonization’ is also represented by changes in neural processing. In Experiments 5 and 6, I found further evidence that cartoony imagery was easier to process than photographic imagery, as measured by the amount of attention – i.e., eye-gaze – that was necessary to respond to cartoony imagery vs. photorealistic imagery. I also found evidence that entirely cartoony displays were more likely to be viewed as congruent when relating symbols to faces compared to mixed media displays. Finally, in Experiments 7 and 8, I found that novel, unknown expressions could be learned easily on both photographic faces as well as cartoony faces, although there was no habituation to cartoony faces while there was to photographic faces. My research demonstrates that cartoony imagery is easier to process compared to photorealistic imagery, and that the extent of this has never fully been described. My research also demonstrates several examples of how cartoony faces show different patterns of allocated attention and different patterns of elicited ERPs compared to photorealistic images.
Lay Summary
People today are surrounded by emojis, children’s cartoons, and comics. Before my research, many assumed that a cartoony face was just a simple version of a real face. However, this assumption is not necessarily correct. I found that cartoony faces are a lot easier to understand and see the expressions on compared to photos of faces. Another thing I found is that people look less at cartoony faces than photorealistic faces – probably because it takes more time to understand a photographic expression. Finally, I found that, although you can learn new expressions easily for both cartoony faces and for real faces, the brain may have a different pattern of processing for cartoony faces than photorealistic faces, so that cartoony faces may not be ‘seen’ as face-like to your brain as real faces are. Overall, my research suggests that cartoony faces are popular because they are so easy to understand.
Preface

This dissertation is an original intellectual product of the author, Lia N. Kendall. All studies were approved by the Behavioural Research Ethics Board (BREB) of the University of British Columbia.

For all of the studies in this dissertation, Lia Kendall formed the research questions, designed the experiments, tested participants, and wrote the manuscripts. Alan Kingstone and Rebecca Todd contributed to the experimental design of all the studies, and contributed edits to the writing process of all the studies as well. For Experiments 1-7, Quentin Raffaelli provided support by running participants and organizing data.

Chapter 3 has been published as: Kendall, L. N., Raffaelli, Q., Kingstone, A., & Todd, R.M. (2016). Iconic Faces are not real faces. Cognitive Research: Principles and Implications.

Chapter 4 has portions which have been submitted for publications as part of: Kendall, L.N., Kingstone, A., Todd, R.M., & Cohn, N. (Under Review). Show me how you feel. Chapter in Iconicity in Language and Literature
# Table of Contents

Abstract ......................................................................................................................... iii  
Lay Summary .................................................................................................................. iv  
Preface ............................................................................................................................ v  
Table of Contents ........................................................................................................... vi  
List of Tables .................................................................................................................. ix  
List of Figures ............................................................................................................... x  
List of Abbreviations ..................................................................................................... xiii  
Acknowledgements .......................................................................................................... xiv  
Dedication ...................................................................................................................... xv  
Chapter 1: Introduction ................................................................................................. 1  
   Purpose ....................................................................................................................... 1  
      Pierce’s theory of signs and other terminology ......................................................... 3  
      Caricatures ............................................................................................................. 4  
      Cartoony images vs. photorealism ......................................................................... 8  
      Cartoony faces ....................................................................................................... 12  
      Understanding Comics ............................................................................................ 15  
      Upfixes and visual language .................................................................................. 16  
      The popularity of cartoony images ...................................................................... 20  
      Overview of Dissertation ....................................................................................... 21  
Chapter 2: Implicit associations between self and images ........................................ 25  
   Introduction ............................................................................................................... 25  
   Experiment 1 ............................................................................................................. 28  
      Rationale and prediction ....................................................................................... 28  
      Participants ............................................................................................................ 29  
      Procedure .............................................................................................................. 29  
      Results ................................................................................................................... 33  
      Discussion ............................................................................................................. 34  
   Experiment 2 ............................................................................................................. 34  
      Rationale and prediction ....................................................................................... 34  
      Participants ............................................................................................................ 35  
      Procedure .............................................................................................................. 35  
      Stimuli .................................................................................................................... 37  
      Results ................................................................................................................... 38  
      Discussion ............................................................................................................. 39  
   General Discussion ..................................................................................................... 39 
List of Tables

Table 1 ......................................................................................................................43
Table 2 ......................................................................................................................78
Table 3 ......................................................................................................................79
Table 4 ......................................................................................................................80
Table 5 .....................................................................................................................111
Table 6 .....................................................................................................................112
Table 7 .....................................................................................................................151
Table 8 .....................................................................................................................152
Table 9 .....................................................................................................................153
List of Figures

FIGURE 1. A drawn caricature can be easier to recognize than a photograph of the same person. ..................5

FIGURE 2. This Japanese ad is one of hundreds. This one is telling train passengers to be aware of the volume of their headphones for the sake of the comfort of other passengers .............................................9

FIGURE 3. The two stimulus sets used in study 1A. The key-bindings for one of these categories would be reversed halfway through the experiment. There were 10 instances of each type of image .......30

FIGURE 4. Shown here are full sets of drawings from two different participants. Note the consistency of the drawings across instances. Participants mostly made happy faces, while a few created a range of expressions. Finally, note the symbolic ‘surprise’ lines used by the second participant. ..........................................................36

FIGURE 5. This participant varied the base shape of her face in her drawings ...................................................38

FIGURE 6. Participants answered more quickly when self-words were assigned the same key as their photos compared to when assigned with their drawings ..........................................................39

FIGURE 7 An example of the five stimulus sets used and a time course of a single trial. The “Cartoon” and “Mid-cartoon” stimulus sets have less complex features than the “Rotoscoped” and “Mid-rotoscoped” sets, and the “Cartoon” and “Rotoscoped” stimulus sets are higher in contrast than the “Mid-cartoon” and “Mid-rotoscoped” sets. Photos may have other low-level featural differences in addition to contrast and featural complexity, but are used here as a baseline non-schematic condition. ........................................................................................................48

FIGURE 8. An example of each type of stimulus set for each type of emotional expression used .............50

FIGURE 9. Accuracy rate for 5 stimulus categories at each presentation time. Error bars represent standard error of the mean. Dotted lines denote low featural complexity stimuli, with solid lines as high featural complexity stimuli. Circle markers denote high contrast stimuli, with triangles as low contrast stimuli. Photo, the baseline stimulus set, in marked in black. Chance is at 25% ................................................................................................................................................56

FIGURE 10. Accuracy for all but the photorealistic photo sets, arranged to illustrate the separate effects of contrast and featural complexity. Error bars represent standard error of the mean. ........................................................................................................................................58
Figure 11. N170 data at sensors P9 and P10. The waveforms on the left show the three stimulus sets averaged across all participants. The charts on the right represent the latencies and magnitude of those individual peaks averaged together; for latency, the photo baseline stimulus N170 was slower than the cartoon and rotoscopied sets; for magnitude, the cartoon stimulus set N170 was greater than the rotoscoped and photo sets.

Figure 12. All the upfix stimuli used in experiment 5.

Figure 13. All the types of upfix/faces dyads, manipulating both the cartooniness of the separate parts as well as their emotional congruency.

Figure 14. A typical trial. Participants were shown each image until response.

Figure 15. Accuracy and reaction times by congruency. Incongruent trials were responded to less accurately and slower.

Figure 16. The different dyad types and their fixation/dwell time ratios. In order: face-photo/upfix-photo, face-photo/upfix-cartoon, face-cartoon/upfix-photo, face-cartoon/upfix-cartoon. Red is dwell time, and pink is fixations.

Figure 17. The different dyad types and their entry times. In order: face-photo/upfix-photo, face-photo/upfix-cartoon, face-cartoon/upfix-photo, face-cartoon/upfix-cartoon.

Figure 18. Example of all the different emotional content that could be in the stimuli from experiment 6.

Figure 19. Example of incongruency distance. The numbers above the heads show what the incongruency distances would be, in this case holding the upfixes constant (at neutral).

Figure 20. The response times for each incongruency distance.

Figure 21. Likelihood to respond “incongruent” based on whether the dyads had two cartoony elements or a photographic face and a cartoony upfix.

Figure 22. Dwell time and fixations by incongruency distance. Having a photo face in a dyad led to more dwell time and fixations to faces, and this was exaggerated for the most incongruent trials.

Figure 23. The cartoony faces that were used in experiment 7 and 8. Top row: icons. Bottom row: symbols.
FIGURE 24. EEG was collected in the green conditions. Participants responded to each face as Sad, Happy, Neutral, or Not an Emotion. After training, all faces were assigned and so there were no more “not an emotion” faces.

FIGURE 25. The novel expressions (in blue) initially elicited slower responses than familiar expressions, but reached the same latency range as the iconic expressions within three blocks. Accuracy only differed in block 1.

FIGURE 26. The magnitude for symbolic (novel) faces increased due to training (in blue). The N170 magnitude for iconic faces (in red) remained constant.

FIGURE 27. The photographic faces used in Experiment 8 to be analogues to the cartoony faces.

FIGURE 28. Accuracies and response times for symbolic faces took a few blocks to converge with iconic faces. While both types of icons were responded to quickly and accurately to begin with, it took roughly 3 blocks for cartoony symbols to catch up. Photographic symbol faces were never responded to as quickly.

FIGURE 29. Cartoony faces showed less of a N170 decrease pre-post compared to photographic faces. These results are not an effect of P1 differences, as described in the results.

FIGURE 30. On the top, difference scores P1-N170, revealing the same effects as the raw N170 scores. On the left, cartoony vs. photographic faces. On the right, iconic vs. symbolic faces. Cartoony faces did not elicit an attenuated N170 pre-post, whereas photographic faces did. On the right, iconic faces showed more of an N170 attenuation pre-post compared to symbolic faces.

FIGURE 31. Symbolic faces showed less of an N170 decrease compared to iconic faces. Again, these results cannot be explained as continuation of P1 patterns.

FIGURE 32. McCLOUD divides artists by image type.
List of Abbreviations

EEG - Electroencephalography

ERP - Event Related Potential

N170 – Negative ERP component at ~170ms found in occipito-temporal locations

P1/P100 – Positive ERP component at ~100ms found in occipital locations
Acknowledgements

I would like to thank the many people who helped this dissertation reach completion. First and foremost, thank you to my supervisors, Beck Todd and Alan Kingstone, who helped me develop my writing and create better experiments. I’d also like to thank my friends and peers who contributed to my experimental design and analyses: Kim Meier, Grace Truong, Kevin Roberts, and Trish Varao-Sousa.
Dedication

To Pusheen, whose face always brought me back to the subject at hand.
Chapter 1: Introduction

Purpose
The simplified, non-realistic face has become ubiquitous. It can be seen when recovering money from an ATM, in children’s movies, on cell-phone apps, in comic books, on buttons, on robots, in medical pamphlets, in textbooks, and a myriad of other locations. The sheer popularity of using such simple non-realistic faces in favour of photographic faces clearly inspires the question: is there some quality that simplified non-realistic faces have that might help explain their widespread use?

This is more difficult to answer than it may seem. Take the common happy face symbol: 😊. What is it? There are many possible easy, straightforward explanations that come to mind. For instance, this face may be described as a cartoon, a simplified face used in children’s media, popular because it is easier to process for children (Mondloch, Le Grand, & Maurer, 2002). Or it could be an emoji, one of the basic emotive faces used in online communication to stand in for non-verbal communication, and popular because of a need for nonverbal, expressive information in texting (Novak, Smailović, Sluban, & Mozetič, 2015). Or, it could be seen as a symbol, a piece of linguistic information which adds tone to a sentence without representing a ‘face’ except in the abstract (Miller et al., 2016).

It quickly becomes apparent that even this single instance of a cartoony image can be difficult to classify. Can the simplified face be a cartoon, and an emoji, and a symbol all at once? Or even an emoticon, a glyph, a pictogram? And do these different terms really describe different underlying perceptual realities of how we perceive the simplified, non-realistic face that would help explain the ubiquity and popularity of the simplified face as a whole?

Rather than attempt to use each of these words and note their individual meanings multiple times, and rather than risk connotations associated with the word ‘cartoon,’ (e.g., to
some it may bring to mind children’s cartoons), this paper will use ‘cartoony,’ to refer to any image which is not photo-realistic in rendering, but is still meant to represent some shared aspect of internal or external reality.

However, this still leaves the original question: do cartoony faces and photographic faces differ in some set of perceptual qualities which could help explain the widespread use and popularity of cartoony faces? This paper will attempt to answer this fundamental question about cartoony imagery from several different directions.

First, in Chapter 2, inspired by the thoughts of comic book artists themselves, I look for evidence that cartoony faces differ from photographic faces due to readers associating them with themselves – or ‘projecting’ themselves into the character due to its simple cartoony face. In Chapter 3, I then seek to address an alternative possibility: that there is a perceptual advantage for cartoony images that facilitates communication or recognition compared to more photorealistic imagery. Expanding on results from Chapter 3, and based on the ideas that cartoony imagery is ‘bound’ together into linguistic relationships, in Chapter 4, I test whether cartoony symbols and cartoony faces are easier to understand together compared to photorealistic symbols and faces. Finally, in Chapter 5, I test whether it is easier to learn a novel cartoony expression compared to a novel photorealistic expression, as any ease of learning could help explain why cartoony images are so commonly used.

This paper suggests, if not a complete answer to why cartoony faces are ubiquitous and popular, a partial answer based on the available research: the cartoony face is a communicative symbol, drawing both on real-world visual similarity (iconicity), as well as linguistic elements (symbolism) to improve communication in a myriad of contexts.
Pierce’s theory of signs and other terminology

As may have become clear in the previous section, categorizing cartoony images can be difficult due to the wide range of images. To help approach looking at cartoony images in a systematic way, in this paper I will use a framework from semiotics, Pierce’s theory of signs (Atkin, 2013a). Piercian theory divides all images into types of *signs*. A sign is anything that represents something else. This is intentionally vague; if one thing represents some other thing, it is a sign. So, for instance, a photograph is a sign for the object that was photographed, a drawing of a person is a sign for that person, and so on. Pierce put forward the theory that all signs can largely be divided into one of three classes. By dividing images created through human history into these three categories it is possible to better describe what a cartoony image is, despite vastly different contexts. While hieroglyphics and cave paintings may appear different, they are both types of visual signs that can be categorized in a Piercian system.

The first type of sign is an index. An index is a sign that directly correlates with the thing it is a sign for, i.e., the signified. An example of indices might be a sign used to warn that a railway crossing is coming up – every time you see a railway crossing sign, a railway will soon come after. Most indices are not constructed images, however; smoke is an index for fire, for instance. Indices will not be discussed at length in this chapter because their use is less common amongst purposefully created imagery.

The second type of sign is the icon. Iconic signs resemble the thing they are signifying. For instance, a happy face, 😊, visually resembles a real human face, and so is iconic. A drawing of a stick figure, although perhaps a poor representation of the human body, is nonetheless iconic because it visually resembles the shape of a human body. In addition, while there are iconic signs that do not rely on visual resemblance (e.g., saying ‘meow’ could be considered an icon of the sound a cat makes), these are not relevant to cartoony imagery.
The final type of sign is the symbol. A symbolic sign is when the sign and signified share no resemblance, and the meaning is only acquired through learning or culture. Examples of symbols would be the $ sign, which represents money but shares no visual resemblance, or letters and words, for which the meanings are learned through culture (Atkin, 2013b; Short, 2007).

It is worth noting here that there can be symbolic and iconic elements within a single image. For instance, traditionally when a western cartoon character fell dead or unconscious, its eyes would become X’s. As Xs do not resemble eyes, such a face holds symbolic meaning in the eyes while still retaining an overall iconicity in the rest of the face. In many cases, a single feature can be both iconic and symbolic. Another notable example would be the many cartoon characters that have a bow in their hair; while a bow is certainly iconically linked to real life hair bows, in many cartoons a bow is also the only sign that a character is canonically female. The iconic bow is being used as a symbolic sign.

Caricatures

One of the first widespread uses of the cartoony image was the cartoony face, in the form of the caricature: an exaggerated and simplified drawing of a person or a character, not unlike a modern cartoon character, that was often used in newspapers to represent famous people or characters (for instance, see Figure 1). The caricature as a phenomenon is interesting due to an important paradox, and one which helps underscore why a comparison of photographs and cartoony faces can be perplexing. The caricature of a celebrity (even an animal caricature!) can be easier to recognize than an actual photograph of that person (Perkins, 1975), even though, as an exaggerated version of that person, it should be less like the person and harder to recognize. Why it should be easier to recognize a caricature – whether the brain stores face information as an aggregate of its most notable features, or whether when cognitively ‘looking up’ a face in
memory we do so on the basis of specific features – became the basis of continuing debate (Perkins, 1975).

**Figure 1. A drawn caricature can be easier to recognize than a photograph of the same person.**

The phenomenon wherein a caricature can be recognized more easily than a photograph can be difficult to study. For instance, in 1985, researchers found no caricature advantage at all. They tested participants for name recall, face recognition, and reaction time for name-face
verification, and found that photographs showed a performance advantage over caricatures. However, all of the caricature drawings that were used in the study were created by a single artist (Tversky & Baratz, 1985), and so it could be that that artist made inadequate caricatures. In support of this concern, another study found that caricatures do indeed have an advantage over photographs in reaction time to a name-face verification task (Benson & Perrett, 1991). Furthermore, this was true of only some caricatures, as it is possible to ‘over-caricaturize’ an image of a person (Benson & Perrett, 1991). That is, there is an optimal amount of exaggeration to use in caricatures in order for them to have a recognition advantage over photographs.

As research progressed, the caricature began to be used as a research tool to reveal aspects about how face-processing might work in general. For instance, one idea put forth about how studying caricatures might shed light on visual cognition is that the way we see objects and people in the real world is based on 2-D templates that emphasize unique features and edges. These features and edges then suggest the contours of objects to our visual system, and so caricatures become easier to recognize than photos because they are clearer exemplars of the unique features of an image (Cavanaugh, 1995; Sayim & Cavanagh, 2011). Supporting this view is evidence that caricatures can help most in situations where distinguishing between similar stimuli is difficult, but that the caricature advantage is ameliorated for stimuli that have been overlearned (Dror, Stevenage, & Ashworth, 2008).

More intriguingly, recent evidence has suggested that a neural network trained on caricatures may also be better at recognizing the faces that the caricatures were based on, much as the human visual system seems to be better at recognizing faces from caricatures (Neves & Proenca, 2019). This study supports a growing body of evidence that there is an element of face-recognition that relies on recognizing a small set of important features (Chang & Tsao, 2017;
Meyers, Borzello, Freiwald, & Tsao, 2015). These important features are capitalized on by exaggerating them in caricature.

For cartoony images, what the caricature studies seem to suggest is that the cartoony image is a natural extension of how we perceive and categorize the visual world. By exaggerating aspects of an image, we create an image that has underscored the aspects that we need to use most for its recognition. Put this way, the goal of this use of the cartoony image is as an iconic representation of the external visual world, just one which caters more to the priorities of the human visual system than does a photograph, counterintuitive as that may seem.

The fact that our visual systems appear to rely on specific features for face recognition rather than detailed iconicity to the object being represented could also suggest that cartoony images are meant to represent vastly simplified internal ideas of the forms of objects and people, much more than they are meant to reproduce something close to our direct perception of those objects and people. That is to say, cartoony images are meant to be less iconic than they may intuitively appear. For instance, some research has found that people tend to caricaturize naturally in their own drawings of objects from memory (Rosielle & Hite, 2009), and that some of our earliest ‘figurative’ drawings in the Paleolithic also show similar evidence of caricaturizing animal images to make them more prototypical (Cheyne, Meschino, & Smilek, 2009). This makes the issue of the caricature more confusing, as it is difficult to gauge how similar or different a very exaggerated cartoony image is compared to a more photo-realistic image, both in intent and in how we perceive it. To what degree each type of image is meant to resemble an object in the world (i.e., is iconic), and to what degree each is meant to represent internal ideas (i.e., is symbolic) is unknown. To help answer this, we must look beyond the caricature, and into other types of cartoony images.
In Chapter 2, I have participants make drawings of themselves, as a kind of caricature, to see if they associate their drawings more with themselves than photographs. In Chapter 3, I directly compare exaggerated, simplified faces to photorealistic faces, to see if one type of stimulus is easier to discriminate expressions on. These two chapters help answer: to what degree are simplified drawings of faces expressing the same information as photorealistic versions of those faces?

**Cartoony images vs. photorealism**

The difference between a caricature and another type of cartoony image has become blurred as more and more cartoony faces have been designed. If you were to take a ride on a Tokyo train, it might seem very familiar to your experience with any other train ride, but for one thing: cartoony messaging informs the rider of rules at every turn. There are cartoons warning you of the dangers of closing doors, cartoons to represent the train brand itself (e.g., as a mascot), and even cartoons that demonstrate how to be a polite rider – featuring everything from somewhat realistic humans to exaggerated animals (Figure 2). It is clear why a subway or train system can benefit from messaging about social standards, but what is not clear is why the images used are always cartoons. What makes an exaggerated, simplified animal face better at explaining polite behavior than a photograph? This is one instance of the overall question I hope to answer with the research in this dissertation.
Figure 2. This Japanese ad is one of hundreds. This one is telling train passengers to be aware of the volume of their headphones for the sake of the comfort of other passengers.

Just as they are found on trains and in texting apps, cartoony images are common in research as well. Cognitive psychology studies often employ cartoony faces for their ease of construction and modification, with the unspoken assumption that they represent real faces, simply with higher contrast, lower featural complexity, and sometimes exaggerated features. Cartoony faces have also been studied in other research domains, such as within comic books, in education, and for disseminating medical information. For example, they have been said to be used as a blank slate that readers can project themselves onto (McCloud, 1994), used as communicative tools in conversation (Ljubešić & Fišer, 2016), and as part of culturally specific visual media (Boyd, 2010; Walker, 1980). This leads one to question: if cartoony faces are being
used preferentially not only by city designers and media creators, but also by researchers, is the reason that cartoony images are being used because they are more effective than photos in some contexts? This paper explores this idea by comparing photorealistic faces with cartoony faces – comparing them for whether we see them as extensions of ourselves (Chapter 2), for whether we can discriminate expressions more easily on them than on photos (Chapters 3 and 4), and whether we learn new information on them more easily than with photographic faces (Chapter 5).

There has been some research into what makes an effective cartoony image, generally in the context of signage or medical pictogram systems (Green & Myers, 2010; Liu, Wang, Song, & Wu, 2010), that led to results which could help answer whether cartoony images are more useful than photographs for certain purposes. For example, McDougall et al. (1999) found that while less visual complexity led to an easier understanding of images, a quality exemplified by cartoony images rather than photographs, images relied on more than being simple or complex to be understood (McDougall, Curry, & De Bruijn, 1999). Here researchers found that beyond visual complexity, a sign was easier to understand if it was more familiar, more concrete (i.e., a more obvious symbolic metaphor depicted), and more meaningful (e.g., had some iconic link to the real-world meaning); however, all features other than visual complexity mattered less as the symbol’s meaning became learned (McDougall et al., 1999). There are additions to what makes a good symbol or icon still to be found today; for example, one odd result that has been discovered is that more aesthetically pleasing images appear to be easier to detect (Reppa & McDougall, 2015).

At their simplest, cartoony images share, compared to photorealistic depictions, qualities such as high contrast, relatively few and simple features (Marcus, 2003; Näsänen & Ojanpää,
Although it is rarely directly discussed, cartoony images also often have somewhat exaggerated features (Medley, 2010). Using the defining properties of cartoony images listed above as the basis for understanding what a cartoony image is allows a direct comparison with photorealistic images. There is evidence that the simplification and exaggeration of photorealistic features associated with cartoony images can improve memory for those images later (Medley, 2010). In addition, past research has shown that more cartoony images – that is, images with fewer and more simple features than photorealistic images – can be perceived in shorter amounts of time compared to their photorealistic analogues (Kendall, Raffaelli, Kingstone, & Todd, 2016).

Based on what we know of cartoony images, it is reasonable to expect that cartoony images will act as versions of photorealistic images that are more rapidly understood and discriminated between. This naturally leads to their proliferation in contexts where processing advantages hold great importance, such as on road signs or in computer menus. What is less clear is how qualities of cartoony images beyond low-level visual features affect how they are understood. While a photograph can easily be understood as directly representing something the viewer could see in their daily life, a cartoony image can also include metaphors or cultural understanding which is not necessarily tied to visual processing (Ferreira, Noble, & Biddle, 2006; McDougall et al., 1999).

Chapter 3 seeks to answer some of these questions by directly comparing different types of face images against one another. By creating a set of stimuli that ranges from fully photorealistic to fully cartoony, I can explore the effect that changing low-level features on a face has on how easy it is to extract information from that face. However, I also go beyond this, in Chapters 4 and 5, by looking for aspects of cartoony images which do not rely merely on low-
level features to distinguish themselves from their photorealistic counterparts. First, by looking for evidence of a ‘schema’ that helps people understand the relationship between a comic symbol and face, and then by looking at how easy it is to learn a new expression on a cartoony face vs. a photorealistic face.

**Cartoony faces**

Studying cartoony images through faces, caricature or otherwise, has an advantage in that it is well established that faces receive special treatment within our perceptual systems. For example, research has found that we are expertly tuned to recognize human faces and their expressions (Rhodes, Byatt, Michie, & Puce, 2004; Tsao, Freiwald, Tootell, & Livingstone, 2006), and that we prefer looking at photographic faces over many other stimuli (Chien, 2011).

Yet even aside from outright caricatures, representations of faces also include cartoons, sketches, emoticons, etc., that can bear little resemblance to real faces, and these media seemingly have a niche in society that photorealistic stimuli do not fill. A clue to potential differences between cartoony faces and photographic faces comes from the observation that cartoony faces are often used in media created specifically for children. This may be because children rely more on low-level or simplified features characteristic of cartoons to process facial emotion (Gao, Maurer, & Nishimura, 2010). Additionally, individuals on the autistic spectrum are better at reading emotions on cartoon faces than on realistic faces (Rosset et al., 2008). These results might suggest that a cartoony face is most notable for its simplicity.

The idea that a cartoony face is merely a simpler version of a photographic face is an attractive one. Cartoony faces are typically higher in contrast (McCloud, 1994), and research supports the idea that higher contrast facial features garner more attention (Neumann, Spezio,
Piven, & Adolphs, 2006), and that higher contrast also results in better face identification (Halit, de Haan, Schyns, & Johnson, 2006).

However, there is no weight of evidence favoring the idea that cartoony faces are merely simplified photorealistic faces. One study found no difference in N170 response (a face-sensitive EEG signal) between cartoony and photorealistic faces (Kendall et al., 2016), which seems inconsistent with evidence of stronger N170 face responses when face features are less ambiguous and more clearly visible (M Eimer, 2000; Schyns, Petro, & Smith, 2007). Another study looking at emoticons did find a stronger face-sensitive signal than to photorealistic faces in EEG (Churches, Nicholls, Thiessen, Kohler, & Keage, 2014), but also found that emoticons did not show the same pattern of N170 changes for inverting emoticons as for inverting faces. This led the authors to conclude that, while emoticons may have face configuration, their features are not necessarily processed as facial features. The relationship between cartoony faces and photorealistic faces is at best unclear. This gives rise to the question: is the cartoony face a simplified photorealistic face, or a less ‘face-like’ type of stimulus? I provide evidence which clarifies the relationship between how cartoony faces are processed (as measured by ERPs and eye-movements) compared to photorealistic faces in Chapters 3 and 4.

Aside from studies looking at how cartoony images are processed by the visual system, there is some other evidence distinguishing between how participants are affected by photorealistic faces and how they are affected by cartoony faces. For instance, Chen et al. (2010) found that the style of cartoony face one is exposed to can bias subsequent face perceptions. The researchers found that 50 minutes of exposure to anime faces – Japanese cartoons where people are depicted with eyes much larger than could be possible on a real person’s face - led participants to prefer real human faces that had larger than average eyes (Chen, Richard, Nakayama, & Livingstone,
These results can be taken two ways. Clearly, the same perceptual system is dealing with both photorealistic faces and cartoony faces if one can use one type of image to bias people to the other. And yet the anime stimuli also seem to have caused participants to prefer faces that were, in a way, less realistic, and less canonical to the real world due to being biased towards a larger average than the norm for photorealistic faces.

There is also evidence that cartoony images are processed less holistically than photographs of faces (Prazak & Burgund, 2014). That is, cartoony faces may be processed more for their individual features rather than as a whole face. The fact that they are perceived less holistically may underlie children’s preference of cartoony images, as children develop holistic face processing later than feature recognition, and may be looking at faces as a constellation of features to begin with (Mondloch et al., 2002). It may also help explain why cartoony faces are easier to process for people with autism (Rosset et al., 2008), where holistic face processing has been found to be at a deficit compared to recognizing facial features in photographs (Joseph & Tanaka, 2003). Another way that cartoony faces may differ from photorealistic faces is that different cultures have different visual ‘vocabularies’ of cartoony images. This can modify how we see any individual cartoony image, so that whether any individual seems canonical or ‘correct’ to the viewer may be informed by cultural knowledge (Cho et al., 2007; Cohn, Murthy, & Foulsham, 2016a). Put another way, there may be a symbolic element to cartoony images that is not always immediately obvious. Thus, for example, the meaning of large anime eyes may not rely entirely on similarities to photorealistic eyes but also on a symbolic cultural understanding of what those eyes represent. I explore these concepts in Chapter 5, by creating faces with novel symbolic expressions both for photorealistic as well as cartoony stimuli, to compare how people learn and perceive new symbolic information on both types of faces.
It is here that we see a clear overlap between understanding the cartoony face as a face, and understanding it as an artistic creation. To this end, it is worth looking at how artists have looked at cartoony faces in the past, and what they think they are creating when they choose to draw a simplified, exaggerated cartoony face to represent something. The use of cartoony images in comic media or animation may help reveal how they differ from their photorealistic counterparts. In Chapter 2, I approach how comic artists have seen their simple cartoony faces by testing one of the claims that comes from artists: that the reason readers like reading comics with simple cartoony faces is that the simplified features make it easier to associate the character with themselves.

Another reason to look at comic artists for answers about simplified cartoony faces is that there are many types of cartoony imagery which appear to be unique to comic books over other media. For instance, comic media contains conventionalized cartoony symbols (Cohn & Ehly, 2016), large stylistic differences from artist to artist (McCloud, 1994), and visual structures which vary from culture to culture (Cohn, Taylor-weiner, & Grossman, 2005). In addition, as many other types of cartoony media (e.g., emojis) draw from comics for their symbols, understanding the use of cartoony faces within comics may be crucial to understanding them overall.

Understanding Comics

As graphic novels became more popular in the 1980’s, comic artists began to take their medium more seriously as an art-form. What they found when trying to analyze their own art was something somewhat mystifying: comics contained symbolic art and visual expressive forms that had never been defined or categorized. The most thorough of these attempts is *Understanding Comics* (McCloud, 1994), a book that can help illuminate some of the concepts on which the studies in this thesis are based.
McCloud attempted to analyze every aspect of the comic medium, dividing his book into several large sections. These included comic vocabulary (the symbols that are unique to comics), the ‘gutter’ and time frames (how the frames in comics separate out meaning and convey a sense of time), an overview of symbolic representations vs. iconic representations in comics, abstract versus concrete art styles commonly used, how early cartoony images became comics, and even how comic panels seemed to have linguistic qualities to them not unlike sentences (Cohn, 2013b; McCloud, 1994).

For instance – like a sentence, panels must appear in a specific order. But the shape of those panels, and how much action is put into a panel (or lack of action – e.g., a pause), can greatly change the meaning. Some comics might have very detailed art, which McCloud believed suggested a representation of the photorealistic world, or have a simple character in the middle of all that detail that he believed the reader was meant to project their personalities onto.

By showing comic strips altered slightly, over and over, and by removing symbols and adding symbols, McCloud successfully expressed how much about comics we still do not know, and how much of the comic medium, and cartoony media as a whole, seem to be built on linguistic processes that have not yet been studied. Arguably it was McCloud who first put forward succinctly the idea that not only can a comic be formed in a linguistic structure, but that the images within that structure could be symbolic, iconic, and unique to the medium.

Upfixes and visual language

There was originally an obvious gap between the work of comic artists like McCloud and more controlled psychological research, but it has begun to close. Within the last decade, psychological research on comics has increased. Recent work has argued that drawings actually engage similar neurocognitive structures as language does (Cohn 2012, 2013), whereby
schematic form-meaning mappings are stored in long term memory, and then combined in production (Cohn 2012, 2013, Wilson 1988, Wilson and Wilson 1977). This may be the quality of cartoony images that separates them from photographic images: their use as linguistic elements.

Visual images are dominantly iconic, that is, resembling the real world. But they can be further characterized along a scale from highly photorealistic to highly schematic or cartoony (McCloud 1993). Moreover, some evidence suggests that discrete manifestation of these forms are processed differently, particularly with regard to faces. For example, contrast and complexity have been shown to be important aspects of extracting information from faces (Gray et al. 2013, Yue et al. 2011). In addition, the well-known deficit in processing emotions from faces associated with autistic spectrum disorder (ASD) is ameliorated when the target faces are cartoons rather than photographs (Rosset et al. 2008, Rosset et al. 2010). It has been proposed that this is because ASD is often associated with a more featural approach to face stimuli (Faja et al. 2009, Goffaux and Rossion 2006, Joseph and Tanaka 2003), and because cartoon faces are more likely to be processed featurally than holistically (Prazak and Burgund 2014).

A recent paradigm has sought to characterize the cognition of drawings not simply in terms of visual perception and their deviance from photorealism. **Visual Language Theory** (VLT) argues that drawn images rely on similar (neuro)cognitive mechanisms as languages. Thus, drawn images are composed of schematic forms stored in long-term memory as lexical items that can be combined in novel ways (Cohn 2012, 2013). The idea of a “lexical schema” has emerged in recent lexical theories of contemporary linguistics (Jackendoff and Audring 2016). A schema is a declarative pattern stored in memory, possibly with internal variables. For example, the word *dog* is composed of a stored configuration of phonology (/dɔɡ/), morphosyntax (Noun),
and semantics (DOG). This noun is a free morpheme, and thus can stand alone. This is different from a bound morpheme like the plural affix -s (as in dogs), which must attach to a stem (like the noun dog). This bound morpheme also involves a schematic mapping between structures, but leaves open internal variables across phonology (/…z/), morphology ([Noun Noun+Affix]), and meaning ([PLURAL(X)]). Thus, to create the plural word dogs, dog satisfies the noun slot in the plural schema and the corresponding structures create a composite meaning ([PLURAL(DOG)]).

Similar mappings hold in “visual lexical items”, albeit with potentially more internal complexity. Consider a basic cartoony face, which uses a systematic (i.e., regularized and repeatable) pattern within and across drawers, and is composed of a circle for a head, dots for eyes, and a line for a mouth. This configuration is stored in memory as a regularized graphic structure, just as dog is stored as a phonological pattern. However, this configuration is understood as a face because of an iconic mapping (Peirce, 1931) between this graphic form and visual semantic memory—i.e., it is recognized as a face because its visual features map to topologically similar visual features in semantic memory (Willats, 1997). In addition, such a form may have internal variables. For example, the line for a mouth may be curved upward (for a smile), downward (for a frown), or a circle (for an exclamation), to name just three. This abstract form-meaning mapping, including its variable “slots,” constitute a schema in the long-term memory of a person who draws faces in this way.

Thus, these form-meaning mappings provide systematic ways of representing visual elements across levels of complexity, from individual graphemes (lines, shapes), to low-level meanings which can combine to form larger novel representations (eyes, hands), to whole figures (people, objects), and whole scenes. Thus, in contrast to assumptions that drawings connect
directly to visual semantic memory, VLT argues that drawings are mediated by a visual lexicon of stored schematic representations, which in turn link to semantic memory (Cohn 2012, 2013).

While VLT argues that iconic visual lexical items have a schematic combinatorial basis, such schematic principles also extend beyond iconic representations. This is again salient in a difference in visual modality between free morphemes, that can stand alone (like people and objects), and bound morphemes, which must attach to a stem (Cohn 2013, Forceville 2011, Cohn 2018). Bound morphemes include things like motion lines to show the movement of a moving object, hearts replacing eyes to show love or lust, or speech balloons to show speech, among many others (Cohn 2013, Forceville 2011). All of these representations must attach to a stem object, and cannot be “free floating” in a drawing. While this visual vocabulary originates in comics and animation, such signs also appear in other visual media, such as emoji. In addition, such lexical items vary across cultures (Cohn 2013, Cohn and Ehly 2016, Tasić and Stamenković 2018), and their comprehension may be modulated by comic reading experience (Cohn and Maher 2015, Cohn, Murthy, and Foulsham 2016, Newton 1985).

**Upfixes** are an example of particularly salient bound morphemes, which are visual “affixes” that are “up” from a head—such as hearts, gears, or birds floating above a character’s head to convey love, thinking, or dizziness, respectively (Cohn 2013, Cohn, Murthy, and Foulsham 2016). (Note, the idea of ‘verbal’ upfixes, unlike standard linguistic affixes which may be verbal or written, is seldom discussed.) Upfixes have been argued to go beyond just specific memorized items, but rather constitute a productive lexical class. In other words, they are not a set of fixed, memorized conventions, but rather upfixes draw upon a schematized template with “slots” for faces and upfixes (Cohn 2013, 2018). Here, the face is the morphological stem, which can otherwise stand alone, and the upfix is a bound morpheme that
affixes to it. While some upfixes are well entrenched in memory (hearts, gears, birds, etc.), the productivity of this pattern allows for novel forms to also be created (rainbows, diamonds, cutlery, etc.). Such a productive schema allows for unconventional upfixes to be construable, though they are recognized as less comprehensible and consistent in meaning (Cohn, Murthy, and Foulsham 2016).

Because upfixes should invoke a clear lexical schema, and that lexical schema often involves iconic elements along with a face, it makes for an ideal place to examine the tension between cartooniness and photorealism in visual representations. Upfixes not only require accessing the direct meaning of an image, but also the emergent, non-iconic meaning that arises out of the combination of the face and the upfix. Thus, if we replace the face and/or upfix with photographic images, they should readily fill the open slots for “face” or “upfix” in the upfix schema. However, photographic images should provide more strained access to this visual lexical schema, which would be more optimized to connect to schematic visual lexical items (drawings), not purely perceptual information (photos).

I explore how upfixes and faces are perceived when photograpic, cartoony, or mixed, in Chapter 4. By varying whether the upfix/face dyad is comprised entirely of cartoony images or whether the face is a photograph, I test whether there is any evidence for a schema influencing how participants respond to the meaning of upfix/face dyads, and whether it is specific to cartoony stimuli.

The popularity of cartoony images
Comparing psychological face research and what we know about caricatures to the ideas of visual language theory brings us back to my initial motivating question: what quality of cartoony images, whether visual or linguistic, gives them a communicative or perceptual advantage that would separate them from photographic images? This is the central question that
this thesis seeks to answer, and understanding how cartoony and photographic faces differ may also help with understanding the popularity of cartoony images based on those differences.

There are essentially three questions which arise from what we know of cartoony faces compared to photorealistic faces which can help approach that larger, overall question, each of which I will answer through the experiments in this paper. These questions are outlined below.

First, is one fundamental difference between a photograph and a cartoony face a form of simplification achieved through a shift in low-level features or exaggeration, and does this simplification allow for an ease of processing compared to photorealistic analogues? This will be answered in Chapters 2-4.

Second, is the underlying aspect of cartoony images that makes them ubiquitous and popular, as VLT postulates, that they contain communicative information that qualitatively differs from photographic faces? That is, is the linguistic and symbolic meaning behind cartoony faces (and other cartoony images) something that photographs do not contain, that allows for a kind of representation set apart from photographs? I attempt to answer this question in Chapters 4 and 5.

Finally, I consider a combination of the above. Could the ease of processing cartoony images, in comparison to photorealistic images, allow for their wide-spread use in an array of media, from comics to road signage? That is, could it be that because cartoony images are so easy to process, they simply can be coopted for many different uses? I believe that this is the possibility which the data favours, as will become clear from the results of each experiment.

Overview of Dissertation

Although cartoony faces are simple and exaggerated versions of photographic faces, they are also used in a variety of contexts where photographic faces are not used (e.g., often in the
context of aiding communication). The current dissertation lays out an investigation into the role of the cartoony face from a cognitive psychology perspective to help bridge the gap between cognitive psychology and media uses of cartoony faces. Specifically, I investigate whether cartoony faces have qualities that distinguish them from more photorealistic stimuli, and how those qualities might underlie the cartoony face’s widespread use. First, through testing ideas about cartoony faces from outside psychology, such as the idea that we ‘see’ ourselves in them (McCloud, 1994). Secondly, through testing how the low-level features of cartoony faces support a difference in how we perceive cartoony images vs. photorealistic images. I then explore whether cartoony faces and photorealistic faces differ not only in behavioural patterns, but also in how they are processed neurocognitively. Finally, I look for any evidence of qualities other than simpler low-level features that cartoony faces may have that could support their widespread use across media.

Through eight studies, I ask several research questions to probe the qualities of cartoony faces that may support their use as tools of communication, and investigate if these qualities are unique to cartoony faces compared to photographs, and if they are, why that might be. These research questions are:

1. Is there any evidence that we see ourselves as better represented in cartoony faces than in photorealistic ones, as suggested by the popular comic artist and comic theorist Scott McCloud?

   In Experiments 1 and 2, I use the Implicit Association Task (IAT) to look for evidence that participants associate themselves with simple cartoony faces. The IAT is thought to represent how closely two categories are associated with each other for a given individual (Greenwald, Nosek, & Banaji, 2003). Thus, in Study 1, by pairing simple cartoony faces or complex cartoony faces with photos of participants, I examine whether simple or complex
cartoony faces seem more ‘self-like’ to participants. In Study 2, I compare the associations between self-words and either photos of the participants or drawings of the participants. In this way, I can determine whether participants associate cartoony images more with themselves.

2. Are the expressions on cartoony faces processed more easily than photographic faces, and if so, to what degree?

   I directly compare different types of face stimuli ranging from fully cartoony to fully photographic in Experiments 3 and 4, by testing the threshold at which an expression can be discriminated behaviorally (Experiment 3), and then finding neural correlates of my behavioral results using EEG (Experiment 4).

3. Is the use of ‘upfixes,’ (i.e., cartoony symbols which are commonly found around cartoony faces in comics), representative of a type of visual communication unique to cartoony images, as has been suggested by comic researchers (e.g., Cohn, Murthy, & Foulsham, 2016)? Or is the widespread use of upfixes simply rooted in the fact that cartoony images are easier to process?

   In Experiments 5 and 6, I compare how both traditional upfixes as well as novel ambiguous upfix stimuli are attended to, using eye-tracking to compare upfix/face dyad stimuli in situations where the entire dyad is cartoony versus when it is not. If upfixes represent a type of visual communication that is unique to cartoony stimuli, the inclusion of photographic elements should make matching the meaning of upfixes to faces more difficult.

4. Are cartoony faces unique in their ability to take on symbolic elements or can novel facial expressions be learned regardless of whether the medium is photorealistic or cartoony?

   In addition to upfixes, which are symbols around the face, there are also many examples of cartoony faces where symbols have replaced iconic elements of the face. For instance, a cartoony face might have $ symbols for eyes, representing greed, or an X for a mouth,
representing a muted face. Past studies have suggested that the N170 ERP is sensitive to face-stimuli and it can be used to discriminate amongst less face-like and more face-like stimuli (Bentin, Allison, Puce, Perez, & McCarthy, 1996).

There are therefore two possibilities. First, that iconic cartoony faces (i.e., faces that visually resemble actual faces) are more ‘face-like’ than symbolic cartoony faces (i.e., faces that require cultural learning to understand) – and therefore evoke a larger N170. The second possibility is that associating new (symbolic) expressions is not unique to cartoony faces, but can be achieved for both cartoony and photorealistic faces.

In Experiment 7, I investigate whether a cartoony face containing a novel symbolic depiction of emotion can evoke patterns of face-sensitive brain activity in the same way as an iconic cartoony face after the meaning of the symbol is revealed to a person. In Experiment 8, I attempt to replicate the results from Experiment 7 and extend my results to photographs, for a direct comparison of how symbolic information in photographs and cartoony images is perceived.
Chapter 2: Implicit associations between self and images

Introduction

In Chapter 1, I put forward a question: what perceptual and communicative qualities separate cartoony faces and photographic faces? One repeated answer comes from comic book artists themselves, in a discussion of different types of images in *Understanding Comics*. In his seminal book, Scott McCloud puts forward the idea that simpler cartoony faces are popular because their simplicity allows readers to project themselves onto the comparatively featureless cartoon face more easily than they could with a complex cartoony face or a photorealistic face (McCloud, 1994). This idea that we ‘see’ ourselves in cartoon images due to their simplicity is contrasted against another popular theory usually found within academic papers: that cartoony faces are popular due to having exaggerated or simplified features that are easy to discern (Mauro & Kubovy, 1992).

The idea that we see ourselves in cartoons or comics is not a new one. As more types of media have been created and consumed (e.g., comics, movies, YouTube videos…), concerns about how people are represented on the screen or in comic books have arisen (Leavitt, Covarrubias, Perez, & Fryberg, 2015). Usually, this has taken the form of how a demographic is portrayed, for instance, if a character is a minority, what qualities are typically applied to them? If a woman is shown, how is she portrayed? Concerns such as these are based on the idea that the viewer will apply the qualities and limitations shown for that character onto themselves (e.g., Merskin, 2008).

There is some evidence that supports the idea that we view cartoons as, at the very least, extensions of ourselves. In one exploratory study, participants were asked how they chose which cartoony images to use in their chat programs. Participants gave several responses, including the
use of cartoons to represent their emotions, to replace non-verbal communication, and perhaps most critically, to represent themselves or some aspect of their personality (J. Y. Lee, Oh, Hong, Lee, & Kim, 2016). Further evidence that we may associate media representations with ourselves comes from the ‘proteus’ effect, which describes how users are influenced by their (typically cartoony) avatar. Specifically, user can behave in line with how they perceive their avatar to be. For instance, being represented by a kind looking avatar may lead a user to behave more kindly (J.-E. R. Lee, Nass, & Bailenson, 2014; Yee & Bailenson, 2007, 2009).

An alternate view to the idea that we see ourselves in cartoony media would be that cartoony images are advantageous for the purpose of communication. That is, cartoony images are simply an efficient way to communicate personality or emotions without associating oneself with the image. There is a wealth of evidence that cartoony images are used for communicative purposes (Henderson, McCulloch, & Herbert, 2003; Medley, 2010; Rosset et al., 2008), and that they may have advantages for certain types of communication such as expressions (Prazak & Burgund, 2014), internal states (Mansoor & Dowse, 2004), or abstract concepts (Naylor & Keogh, 2013). This alternative interpretation would propose that, rather than associating ourselves with cartoony images, the cartoony images communicate information to us that we may then apply to ourselves (and, images which are easier to understand are then easier to apply to oneself). This would differ from the proteus effect in that the information people would apply to themselves is explicit and conditional – people would not see an image as themselves, but as suggesting information they could accept or reject. This is idea is also more in line with evidence that shows that caricatures of a person can be easier to recognize than photographic images – the exaggeration of the caricature, a type of cartoony image, is being used to communicate the most important facial information to the viewer (Mauro & Kubovy, 1992; Medley, 2010).
This leads to the question of whether when we see a cartoony face we see it as a representation of ourselves, or if we see it as a medium for communicating some type of information. When participants say that they use a cartoony image in an online messaging app to ‘represent their personality,’ does that mean, for example, that the cartoony cat that they posted is being used as a version of themselves, or rather as a method to easily display the expression the cat is exhibiting?

One way to approach understanding how people relate their ideas of self to simple cartoony images vs. more complex images or photorealistic images is provided by the Implicit Association Task (IAT). The IAT was developed as a measure of unconscious bias. It requires participants to classify items by pressing designated response keys, such that faster response times for a paired set of categories is interpreted as indicating a stronger implicit association between those categories (e.g., Canadian and good) (Greenwald et al., 2003). This paradigm has been widely applied across different domains of psychology research, and has been used with different types of stimuli, including words, images, and sounds (Farmer, Maister, & Tsakiris, 2014; Greenwald et al., 2003; Ma & Han, 2010; Pinter & Greenwald, 2004).

In Experiment 1, I used the IAT to determine whether people associated photos of themselves with more simply rendered cartoony faces versus more complexly rendered cartoony faces. In this way, I could test the validity of the idea that people are more likely to see themselves in simple cartoony stimuli. If people associate simple cartoony images with themselves, they should be able to pair simple cartoony faces with photos of themselves more easily than they would pair complex cartoony faces with photos of themselves. In Experiment 2, I sought a stronger effect by looking at what kind of depiction people would associate themselves with more: photos of themselves, or drawings of themselves. This allowed me to
compare two sets of stimuli that both explicitly represented the participant, differing only in how they were created. If people associate themselves with simple depictions of themselves more than photorealistic depictions of themselves, the IAT should reveal that they respond to their drawings with self-words more easily than they would pair their own photographic faces with self-words.

Experiment 1

Rationale and prediction
Experiment 1 drew inspiration from *Understanding Comics*, where author McCloud posited that simple cartoony faces are different from more complex faces in that people ‘see’ themselves in the simplified face (McCloud, 1994). When searching for a way to compare the qualities of cartoony and photorealistic faces, this seemed like a good place to start. I wondered: could McCloud be right?

I predicted that if people see themselves in simple cartoony faces, I would find a strong effect using an IAT showing that people found very simplified cartoony faces to be more associated with themselves than complex cartoony faces. In terms of the IAT, this would mean the simple cartoony faces assigned to the same key as my participants’ photos would elicit lower reaction times than in conditions where complex faces were assigned to the same key as my participants’ photos. Alternatively, if the complex faces elicited lower reaction times, it would mean that complex faces and the complex photographs of my participants were easier to respond to when assigned to the same key. Finally, if neither complex nor simple cartoony faces elicited lower reaction times when assigned to the same key as participants’ faces, it would suggest that neither type of image was strongly associated with the participants in their minds.
Participants

27 Participants (22 identified as women, Mean age= 19.82, 5 as men, Mean age= 20.8) participated for course credit. A sample size of 25 participants was determined prior to data collection as sufficient for the intended analyses based on past research from our research group. Two additional participants were tested as data collection continued for the entire week within which the 25th participant was collected.

A post-hoc power analysis of these data revealed that this number of participants was sufficient to reach a power of 70%, or 70% likely to find an effect if there was one, assuming an effect of a moderate size (.5) and a two-tailed alpha of .05.

Procedure

Participants performed the entire experiment in a dimly lit room, with images presented on a neutral grey background (125, 125, 125 on an RGB scale). They responded to each type of stimulus using the ‘z’ and ‘m’ keys. These key bindings were counterbalanced, and the phases described below were counterbalanced across participants. Each image was 250 x 250 pixels, and presented in the center of the screen for a duration randomly selected from 900 to 1500 ms. Following a response there was 500 ms intertrial interval. Blocks 1, 2 3, 5, and 6 (the training blocks) all had 20 trials each. The test blocks, 4 and 7, each had 40 trials. This resulted in 180 trials in total.

The IAT procedure was adapted from Greenwald et al. (2003), and used the typical 7 stage procedure outlined by their research group. At the beginning of each experiment, photos were taken of the participant as well as of a friend they brought into the experiment with them, and each these two categories of photos were assigned to separate keys. The remaining
categories were very simple cartoony faces and complex cartoony faces (gender matched to the participant).

**Figure 3.** The two stimulus sets used in study 1A. The key-bindings for one of these categories would be reversed halfway through the experiment. There were 10 instances of each type of image.

In the IAT, participants classify four categories of stimuli using two response keys, so that two categories are mapped to each key. In this experiment, participants classified ‘self-photos’ (photos of the participant), ‘other photos’ (photos of a second participant who accompanied the first participant), simple cartoony faces, and complex cartoony faces. Self-photos/other photos were never responded to using the same key, nor were simple cartoony images/complex cartoony images. See Figure 3 for stimulus examples.

In total, the IAT involves 7 phases, but only phases 4 and 7 are test phases.
In phase 1, Participants began the experiment by training themselves to classify self-photos with one response key and other photos with the other response key. In phase 2, participants were trained classifying simple cartoony images with one key, and then complex cartoony images with the other response key. Phase 3 was a training phase where all four categories were responded to, so that, for instance, both ‘self’ and ‘simple’ responses were mapped to ‘z’, and ‘other’ and ‘complex’ responses were mapped to ‘m’. This training phase was then repeated as a test phase (Phase 4).

In Phase 5, participants were “retrained.” Two of the stimulus categories now had their response key configurations reversed. So, for example, where previously simple/complex were responded to using ‘z’ and ‘m’, respectively, ‘simple’ would now be responded to using ‘m’ and ‘complex’ would be responded to using ‘z’. Phase 5 had participants practice this new configuration. The second retraining phase, Phase 6, then contained all four categories of stimuli again, so that both ‘self’ and ‘complex’ would be responded to using ‘z’, and ‘other’ and ‘simple’ would be responded to using ‘m’. This retraining phase was repeated as a test phase (Phase 7).
Out of all of the phases described above, there are then only two test phases: one where the ‘z’ key is used to respond to self/simple and the ‘m’ key is used to respond to other/complex, and then a second phase where the ‘z’ key is used to respond to self/complex and the ‘m’ key is used to respond to other/simple. Response speed can be taken to indicate the degree to which two categories are associated (Pinter & Greenwald, 2004), and so the degree to which participants were faster to respond to self/simple together compared to self/complex could be understood as a stronger association between these categories.

There are many ways to score the IAT. For instance, it is possible to use blocks 3 and 4, as well as 6 and 7, rather than simply 4 and 7. However, blocks 3 and 6 are still considered ‘training’ blocks, and typically have longer response times than the ‘test’ blocks of 4 and 7. In

<table>
<thead>
<tr>
<th>Phase</th>
<th>Key 1</th>
<th>Key 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Self-photos</td>
<td>Other photos</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>cartoony images</td>
<td>cartoony images</td>
</tr>
<tr>
<td>Phase 3</td>
<td>Self/Simple</td>
<td>Other/Complex</td>
</tr>
<tr>
<td>Phase 4 (test)</td>
<td>Self/Simple</td>
<td>Other/Complex</td>
</tr>
<tr>
<td>Phase 5 (retrain)</td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>cartoony images</td>
<td>cartoony images</td>
</tr>
<tr>
<td>Phase 6</td>
<td>Self/Complex</td>
<td>Other/Simple</td>
</tr>
<tr>
<td>Phase 7 (test)</td>
<td>Self/Complex</td>
<td>Other/Simple</td>
</tr>
</tbody>
</table>

**Table 1.** Participants are trained with two categories mapped onto one response key, and another two categories onto a second response key. After a test phase, two categories have their key configuration reversed. Note that these key-bindings were counterbalanced (e.g., in half of participants Self/Complex and Other/Simple were the initial mappings). Grey rows represent test phases.
block 3 as well as 6, participants are still learning to assign two sets of stimuli to a single key. I believed a training effect could contaminate my data, which led me to using only blocks 4 and 7.

The IAT can also be scored using block median, reciprocal latency (1000/latency), log-transformed latencies, as well as other values (Greenwald et al., 2003). Ultimately, I decided on means so as to provide a consistent narrative with the analysis choices across the experiments in this thesis. This also provides a simple understanding of IAT effects consistent with other cognitive experiments that do not rely on unusual statistical transformations to determine positive effects.

Results

Response times were extracted for the trials in both blocks 4 and 7 for each participant, and then compared for trials where the self-photos were assigned to the same key as photos of the participant. These trials were averaged and compared using two-tailed paired t-tests.

The response times for trial blocks that paired photos of participants with simple cartoony faces were not significantly faster than trial blocks where photos of participants were paired with complex cartoony faces \((p=.22)(543\text{ vs }570\text{ms})(\text{see Table 1})\).

To confirm that this manipulation failed to show a difference in implicit associations, the data were also examined using a Bayes factor (null/alternative) (Jarosz & Wiley, 2014). This analysis returned a \(\text{BF}_{01}=1.411\), suggesting that the data were 1.411:1 in favour of the null hypothesis, or that it was not easier or harder for participants to respond when self-photos were assigned to the same key as complex images as when assigned to the same key as simple images (Jarosz & Wiley, 2014). While this is considered a weak effect in Bayesian statistics, it nonetheless provides positive evidence that neither of the cartoony image conditions were more associated with photos of the participant as measured by the IAT.
An additional analysis was performed, using only ‘other’ photographic faces, comparing trials where other-faces were assigned to the same key as simple cartoony faces against trials where other-faces were assigned to the same key as complex cartoony faces. This analysis also found no difference between conditions (p=.087). This secondary analysis confirms that there was no RT facilitation due to complex photographic stimuli being paired with complex cartoony stimuli.

Discussion
This experiment compared how much participants associated photos of themselves with simple cartoony faces vs. complex cartoony faces, based on the idea that people may associate themselves more with more simplified faces. This is not what was found – Associations between complex cartoony faces and photos of participants were not significantly different from associations between simple cartoony faces and photos of participants. I failed to disconfirm the null hypothesis that complex cartoony faces and simple cartoony faces are equally unrelated to participant’s ideas of self.

Experiment 2

Rationale and prediction
Results of Bayesian analysis for Experiment 1 indicated that it was likely that the null effect was reliable, and that participants were not associating simple cartoony faces with themselves any more than complex cartoony faces. However, one explanation of this null finding is that participants were perceiving both the simple and complex cartoony faces as representing ‘other’ people. To remedy this, I designed Experiment 2 to compare simple cartoony versions of the participants vs. photographic versions of themselves vis-a-vis which would be more associated with self-related words.
Based on the results from Experiment 1, I predicted that participants would associate photographs of themselves more with their concept of ‘self’ than cartoony versions. This would be reflected in shorter reaction times for conditions where photographs were assigned to the same key as self-words compared to conditions where cartoony drawings of the participants were assigned to the same key as self-words.

Participants

27 Participants (22 identified as women, Mean age= 19.82, 5 as men, Mean age= 20.8) participated for course credit. As before, a sample size of 25 participants was determined prior to data collection as sufficient for our intended analyses. Data collection ceased at the end of the week after data from the 25th participant was obtained. For this experiment, 1 participant was excluded because of a task error, leaving 26 participants. Although the previous study found no effect with a similar number of participants, I did not believe that it was due to a lack of power.

Procedure

As with Experiment 1, Experiment 2 used the IAT. In this experiment, however, the categories were: photos of self/cartoony images of self, and self-words/other-words. Before each experiment, participants had their photos taken, and also were told to draw ten versions of themselves in a computer program (see examples in Figure 3). Drawings were completed on 250 x 250 pixel templates in Paint.NET (freeware: https://www.getpaint.net/). Photos taken of the participants versus drawings the participants had drawn of themselves were one set of stimuli. The other two stimulus categories were self-words such as “I,” “me,” and “mine,” and other-words such as “you,” “them,” “hers.”
Each participant underwent the same 7 phases of the IAT as were described in Experiment 1, where self-words/self-photos and other-words/self-drawings were classified using two keys (Phases 1-4). Afterwards, the stimulus categories reversed, so that self-words/self-drawings and other-words/self-photos were classified together (Phases 5-7). In half of participants, the order of these associations was reversed, so that self-words/self-drawings and other-words/self-photos were the key mappings for Phases 1-4.

Using the IAT, one might discover which categories participants more readily associated: photos of themselves and self-words, or drawings of themselves and self-words. By looking at
response times for the self-words/self-drawings mappings and comparing them to self-words/self-photos, one can infer which association is stronger (Greenwald et al., 2003).

Stimuli

The drawing stimuli, as shown in Figures 4 and 5, displayed many interesting qualities worth mentioning. Each participant was given 1 minute per drawing, and so spent 10 minutes drawing 10 faces to represent themselves. The only instructions they were given were to “draw an image of yourself within this square, and within a minute.” Participants were not given any visual aids to help them draw; however, they completed these drawings directly after seeing the photos of themselves.

Descriptively, a number of patterns were observed. Out of 27 participants, only 2 varied the shape of their drawn faces. A different 2 participants created an image that was not ‘forward-facing’. Four participants used conventionalized symbols from comics or cartoons in their drawings: e.g., lines extending from the head for surprise, sweat drops flying from the head, or hearts for eyes. However, other ‘iconic’ aspects of cartoony images which could be considered conventionalized symbols learned from comics or animation were also present: specifically, exaggerated ‘blush’ marks or mascara for beauty, and the tongue-sticking out to express silliness.

Taking the drawings as a whole, 75% were drawn with happy expressions, showing a clear majority. After happy, neutral was the most common expression (10%), followed by silly expressions (6.5%) (e.g., sticking a tongue out or winking).
Figure 5. This participant varied the base shape of her face in her drawings.

Results

Once again, reaction times were averaged for block 4 and block 7, the two test blocks. Responses were averaged for trials where self-words were assigned the same key as drawings of the participant or with photos of the participant, so that those blocks could be compared. Trials were analyzed using two-tailed paired t-tests.

There was a significant difference between the two association blocks for response times, \([t(25)= 2.88, p= .008, d= .564]\). Response times were faster when self-words and photos were assigned to the same key than when drawings and self-words were assigned to the same key (see Figure 6).
Discussion

This result is the opposite of what we should observe if people saw themselves in simple drawings of themselves. Instead, the results of study 2 suggest that participants associated photorealistic images of themselves more with ‘self’ than drawings they themselves had created to represent themselves. This does not support the idea that people project themselves onto simple cartoony images, or see themselves in such images. However, it is possible that the drawings participants made of themselves did not resemble their photographs enough to create an association between ‘self’ and the drawings, despite the fact that participants knew the drawings were meant to be representations of themselves.

General Discussion

In this study, I used the IAT to compare different types of stimuli to address the question of whether participants associate themselves more with simple cartoony faces compared to more complex images. Across two experiments I found either no evidence to support the idea that...
people associate themselves with simple cartoony imagery, or evidence that they see more complex representations of themselves (i.e., photographs) as associated more with their concept of self.

In Experiment 1, I found no evidence that participants associate themselves more with simple cartoony imagery compared to complex cartoony imagery, as well as weak evidence that both types of imagery were equally easy to pair with photos of the participants. Specifically, when photos of participants were assigned the same key as simple cartoony images were compared to those where they were assigned the same key as complex cartoony images, there was no difference in response time. This means there was no evidence from this experiment to support the idea that participants were associating the simple images more with their own image compared to the complex cartoony images. In a Bayesian reanalysis of the data, I also found that the null hypothesis was more likely than the alternative hypothesis, that is, that it was more likely that there were no associations between either type of cartoony image and photos of the participants (Dienes, 2014).

In Experiment 2, I compared blocks of trials where ‘self’ words were assigned the same key as either photos of the participants or cartoony drawings the participants had made of themselves. The results showed that assigning photographs of the participants with ‘self’ words led to faster response times than when the ‘self’ words were assigned to the same key as the participant’s drawings. This suggests that participants associate their own photographic images more with their concept of self, compared to drawings they had made to represent themselves. A caveat to this finding is that my participants may not have had the artistic skill to render themselves accurately enough to match their internal image of themselves – an issue which may
also be behind the lack of caricature effects sometimes found in the literature (Mauro & Kubovy, 1992).

Together, these results contradict the proposal that people see cartoony images as projections of themselves – whether they are simple or more complex. Indeed, there was no difference between the two types of cartoony faces used in Experiment 1 even though the stimulus types were rendered differently. So what can these data tell us about how people view cartoony images?

The IAT I used in this paper has limitations. A common objection to its use is that the IAT does not measure implicit biases, but instead the activation of memory associations people have learned, and that these associations may still be judged as invalid before they become conscious (Gawronski, Lebel, & Peters, 2007). For instance, if people see media that consistently pairs “refreshing” with “Sprite,” they will become more easily activated together in memory, which could then be measured by the IAT – however, if you hate the taste of Sprite, this does not mean you will actually hold this bias consciously. This account helps explain why priming participants with a viewpoint (e.g., “Which sodas do you find refreshing?”) may increase the IAT effect (Olson & Fazio, 2003).

What this means for cartoony imagery in this chapter is that my data suggest that there is no unconscious memory structure linking simple cartoony imagery with people’s concept of ‘self,’ even in situations where people purposefully create a depiction of themselves, as in Experiment 2. People associate themselves more with photographic representations of themselves than with their own drawn depictions– likely because photographs (and mirrors) are the representation that they have been exposed to as representing themselves more often. While it may be possible to build an association between a cartoony image and a person’s identity
purposefully which could then be measured using the IAT, based on the data presented in this chapter it is unlikely that people naturally view cartoony images in such a manner.

Ultimately, this chapter found no evidence for participants associating themselves with cartoony imagery, or projecting themselves into simple cartoony faces. The idea that people see themselves in cartoony imagery may be an artifact of how they view cartoony imagery in the first place. If there is an image that is easier to process due its low-level features (Prazak & Burgund, 2014), and if that image is easier to manipulate due to its simplicity so that information can be prioritized (W. Yang, Toyoura, Xu, Ohnuma, & Mao, 2016), it may also communicate information to the viewer in such a way that they more readily identify with it, giving rise to the spurious idea that people ‘see’ themselves in the image. Rather than searching further afield for evidence that people associate themselves with simple cartoony imagery, it is useful to first examine the alternative idea that cartoony images hold communicative advantages primarily due to their low-level features.
### Table 1

<table>
<thead>
<tr>
<th></th>
<th>RTs</th>
<th>Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Ex. 1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple</td>
<td>543</td>
<td>15</td>
</tr>
<tr>
<td>images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex. 1.</td>
<td>570</td>
<td>18</td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex. 2.</td>
<td>596</td>
<td>17</td>
</tr>
<tr>
<td>Photos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex. 2.</td>
<td>642</td>
<td>22</td>
</tr>
<tr>
<td>Drawings</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3: Effects of schematization on image perception

Introduction

Chapter 2 suggested that there is no strong evidence for people viewing simple cartoon faces as themselves, or ‘projecting’ into them. However, cartoony faces remain very popular – so what are their advantages? McCloud (1994) underscored how cartoony images are exaggerated and simpler versions of more detailed faces, and in the psychology literature, ‘schematic’ faces are used for that very purpose. Where the literature and the artist’s theory intersect is simplicity: I propose that iconic faces use simplified and enhanced visual features to facilitate the communication of emotion.

Evidence that low-level visual features such as contrast and complexity influence identification of both facial identity and expression is consistent with this view. A large body of previous research suggests that face perception is heavily influenced by differences in stimulus type, especially in low-level visual features (Crouzet & Thorpe, 2011; Goffaux & Rossion, 2006; Sung et al., 2011; Yue, Cassidy, Devaney, Holt, & Tootell, 2011). For instance, contrast can provide an advantage in face identification (Halit et al., 2006), and high-contrast facial features elicit longer fixations than lower-contrast features (Neumann, Spezio, Piven, & Adolphs, 2006). Moreover, both contrast and spatial frequency profiles have been found to facilitate identification of fearful faces (Gray, Adams, Hedger, Newton, & Garner, 2013; Yang, Zald, & Blake, 2007).

If facilitated processing of iconic images is indeed predicted by low-level features such as contrast and simplicity, underlying differences in cortical processing should be reflected in event-related potentials (ERPs). The P1 is an early perceptual ERP sensitive to low-level features in its latency and amplitude (Kappenman & Luck, 2012; Woodman, 2010). It is delayed by decreasing the luminance of a stimulus (Halliday, McDonald, & Johnson, 1973; Fimreite,
Ciuffreda, & Yadav, 2015); it is delayed and lower in amplitude when a stimulus has lower contrast (MacKay & Jeffreys, 1973; Hosseinmenni, Talebnejad, Jafarzadehpur, Mirzajani, & Osroosh, 2015); and it is lower in amplitude for smaller relative to larger stimuli (Asselman, Chadwick, & Marsden, 1975). Early studies also found that stimuli with higher levels of pattern detail (i.e., finer checks on a checkerboard pattern) evoke larger P1 amplitudes than stimuli with larger low-level features, indicating a smaller amplitude P1 with reduced complexity (Oken, Chiappa, & Gill, 1987; Leserve & Romand, 1972; Zaher, 2012). And disorders that negatively impact low-level visual processing, such as multiple sclerosis, are associated with delayed P1 components (Halliday, McDonald, & Mushin, 1973; Zaher, 2012). Together, these findings suggest that the P1 should be sensitive to clear and unambiguous features on a face. Specifically, as low-level features of an image become more cartoonized – i.e., simpler and higher in contrast — they should evoke a shorter latency and lower amplitude P1.

In contrast to the P1, the N170 is a face-sensitive ERP component modulated by emotional expression (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Blau, Maurer, Tottenham, & McCandliss, 2007; Eimer, 2011, Hinojosa, Mercado, & Carretié, 2015). N170 amplitude and latency have been found to be sensitive to spatial frequency information (Halit et al., 2006) as well as contrast (Y. Lu, Wang, Wang, Wang, & Qin, 2014) and image complexity (Churches et al., 2014). Despite these established findings, an open question concerns whether amplitude or latency patterns observed in the P1, due to increased simplification of cartoon images, are carried on to the N170.

Convergent findings suggest that this may be the case. Although the P1 is neither a face-specific nor emotion-sensitive component, there is some evidence that it is the first component that is sensitive to differences in face stimuli in childhood. In contrast, the N170 develops face
sensitivity later in development (Taylor, Batty, & Itier, 2004). These findings support the conclusion that low-level features in faces play a greater role in face discrimination earlier in development, and face-sensitive processing is built on this detection of basic visual features. Such findings are also consistent with the notion that the P1 plays a role in face detection without requiring sensitivity to faces as a specific category. Moreover, unlike the N170 (Desjardins & Segalowitz, 2013; Rossion & Caharel, 2011), the P1 is modulated by the presence or absence of face features even in scrambled faces. Thus, convergent evidence indicates that the P1 reflects cortical activity implicated in early feature detection which contributes to face sensitive processing indexed by the N170. Although the holistic processing that is distinctive of face processing may not occur prior to the N170 (Rossion, 2014), this earlier component may influence later face processing by allowing for faster processing if features have been more quickly recognized.

In sum, previous research suggests that face perception is heavily influenced by differences in stimulus type, especially in low-level features, and that such manipulations modulate rapid cortical activity that precedes holistic face perception. Thus, enhanced communicative capacity through intensification and simplification of low-level features in iconic images such as emoticons and cartoons may underlie their ubiquity, despite their seeming dissimilarity from real-world stimuli. The present study tested the hypothesis that as faces become more iconic, emotional information becomes easier to access. I thus created face stimuli increasing in schematization through simplification (reduced complexity) and enhancement (higher contrast). In Experiment 3, I examined whether schematization affected the discrimination of emotional expressions across a range of presentation times. I predicted that accuracy would increase with schematization, and that this would become more pronounced as
presentation time decreased. In Experiment 4 I used ERPs to examine how neural responses to manipulations of complexity reflected schematization advantages in emotion discrimination. I predicted differences in amplitude and latency with simplification that would correspond with my behavioural results.

Experiment 3

Rationale and prediction

Experiment 3 was designed to directly compare types of face stimuli which have not been compared in the past, specifically, faces that vary in how cartoony they are, ranging from fully cartoony to fully photographic. Past studies reported a floor effect for discriminating (fearful) expressions on masked face stimuli presented at 33ms (Pessoa, Japee, & Ungerleider, 2005); but this was observed only for photographic faces, which can differ from cartoony faces in many ways. This seemingly absolute threshold of expression discrimination for expressions on photographic faces in the past, at 2/60 frames (how 33ms was chosen) provided a good lower presentation rate at which to compare photos to cartoony faces. The final and full set of presentation rates were drawn from 1- 4 frames out of 60 – or 17, 33, 50, and 66 ms presentation times.

I predicted that due to the higher contrast of cartoony faces and their lower featural density, cartoony faces would be more discriminable than photorealistic faces, and that the more cartoony the face the easier it would be to discriminate expressions. I specifically hoped that I would find that cartoony faces could be discriminated at lower than 33ms presentation times, as this would reflect an obvious difference against previous results with photorealistic faces.

Participants

50 undergraduates (36 identified as female, mean age = 20, 14 identified male, mean age = 20.8) participated in return for course credit. A sample size of 25 participants per level for each
between-subject independent variable was determined before data collection as sufficient to find meaningful effects based on previous studies done by our research group. The present number was determined as the experiment included a comparison of mixed vs. blocked trials as the only between-subject factor. This factor was originally included to control for order effects, but as none were found, it will not be discussed further. Data collection ceased when 50 participants had been tested.

Stimuli

**Figure 7** An example of the five stimulus sets used and a time course of a single trial. The “Cartoon” and “Mid-cartoon” stimulus sets have less complex features than the “Rotoscoped” and “Mid-rotoscoped” sets, and the “Cartoon”
AND “ROTOSCOPE” STIMULUS SETS ARE HIGHER IN CONTRAST THAN THE “MID-CARTOON” AND “MID-ROTOSCOPE” SETS. PHOTOS MAY HAVE OTHER LOW-LEVEL FEATURAL DIFFERENCES IN ADDITION TO CONTRAST AND FEATURAL COMPLEXITY, BUT ARE USED HERE AS A BASELINE NON-SCHEMATIC CONDITION.

Figure 8 illustrates the stimuli, which consisted of five categories of faces employing increasing degrees of schematization: cartoon (non-realistic iconic faces where the only features present are used for communication, such as the eyes and mouth), mid-cartoon (the same as cartoon, but with a ‘skin tone’, so that the face has a shade of grey darker than the white background), rotoscoped (photographs that have been schematized by using a technique for drawing over the photograph, creating a heavy outline to emphasize high contrast features like the eyes, mouth, and nose, while removing others), mid-rotoscoped (the same as rotoscoped but leaving the average skin-tone of the photo intact so that the face is darker than the white background), and unmanipulated realistic photos, which acted as a non-schematized control, or baseline stimulus set. Each stimulus group was comprised of an equal number of images.
Figure 8. An example of each type of stimulus set for each type of emotional expression used.
The cartoon and mid-cartoon stimuli were constructed using basic photo manipulation software, GIMP (Kimball & Mattis, 1996). The positions of features were varied slightly from image to image in the mid-cartoon set so that there were several versions of each face (e.g., the eyes could be shifted to be slightly wider, so that not every cartoon face is perfectly identical), and then the ‘skin-tone’ was removed (i.e., turned white) from each of these to create the cartoon stimulus set. To create the rotoscoped and mid-rotoscoped stimulus sets, the photo stimulus set was processed using rotoscoping software by Synthetik (Dalton, 1999). In this way, in addition to the realistic photos, which acted as a non-schematic control, I created 4 stimulus sets which non-linearly spanned a range from realistic to schematic (i.e., cartoon) faces. As a manipulation check, I asked another group of 60 participants to rank order the stimulus sets from least realistic to most realistic. 70% ordered them in the order illustrated above (the second most common configuration being a simple switch of the mid-cartoon and rotoscoped sets, which comprised 20% of responses).

With the photo stimuli excluded, these stimuli can also be grouped by two factors, contrast and featural complexity. The mid-rotoscope stimulus is identical to the rotoscoped stimulus set, but with lower contrast, and the mid-cartoon is identical to the cartoon stimulus set but with lower contrast. Likewise, the cartoon and mid-cartoon stimulus sets can be seen as less featurally complex versions of the rotoscoped and mid-rotoscoped stimulus sets respectively. Each stimulus set included four categories of facial expression: disgusted, happy, surprised, or neutral. These expressions were selected as commonly recognized basic emotions (e.g. Ekman, Sorenson, & Friesen, 1969) that were physically distinct from each other and of mixed valence. Because surprise can be either negative or positive, I used the term “shocked” to emphasize negative valence and reduce participant confusion. There were 8 variants of each expression in
each set, represented by different individuals’ faces in the more realistic sets, and by varied feature positions in the cartoon and mid-cartoon sets. Thus, there were 160 images in total (8 variants x 4 expressions x 5 stimulus sets). For the purpose of rotoscoping, I used faces from a database of emotional expressions created to be used for animation. See Figure 8 for an example of each emotional expression for each stimulus set.

All images were presented using PsychoPy software (Peirce, 2007). Each of the 160 images was shown 4 times, corresponding to 4 possible presentation times. The four presentation times were 16.7, 33.3, 50, and 66.7 ms, which corresponds to 1, 2, 3, and 4 frames on a monitor with a 60hz refresh rate, and were chosen randomly on a trial by trial basis. Therefore, in total, there were 640 trials (160 images x 4 presentation times).

A further manipulation concerned presentation of stimulus type in a blocked (i.e., all cartoon images together, all photos together…) or event-related design (any stimulus type could be chosen for each trial, randomly). However, there was no statistical difference between this between-subjects factor (p=.71), and so it will not be considered further.

Procedure

All participants performed the task on a laptop in a dimly lit testing room. Images were displayed on a neutral grey background (125, 125, 125 in the RGB colour system).

Figure 7 illustrates the sequence of events in a typical trial. Before each trial, a fixation point (+) was positioned on the middle of the screen, and participants were told to keep their eyes on its location. A random line mask would appear for 83 ms (i.e., 5 frames out of 60), then one of the face images at one of the four presentation times, and then another mask, followed by a response screen. Stimuli were forward and backward masked to ensure that they were only visually discernable for precise presentation times. Participants were asked to identify the expressions of the images presented to them, as quickly and as accurately as possible, using
numbered keys: 7, 8, 9 and 0, representing disgusted, happy, shocked, neutral respectively. The next trial was presented following the previous response (i.e., as there was 1000 ms of fixation preceding each stimulus presentation, the ITI was always 1000ms).

Subjects were given a practice session with feedback at the beginning of the experiment to familiarize them with the program before the actual recorded trials. Subjects were also given the option of resting breaks every 128 trials. Accuracy was recorded for each trial.

Results

Figure 9 shows all stimulus sets at each presentation time. A 5x4x4 (level of schematization x expression x presentation time) repeated measures ANOVA was conducted to analyze accuracy data (there were insufficient correct trials in each condition for RT to be a meaningful measure, and so no analyses on RT data are included here). Although expression was included as a factor in analysis, there were no meaningful interactions with presentation time or stimulus type (i.e., all significant results trended towards disgust simply showing a slightly more exaggerated pattern of results, and neutral expressions showing an attenuated pattern of results, while always showing the same order of discriminability for each stimulus type). Thus, expression results will not be reported here; however, all values for all conditions can be found in Table 2.

All contrasts were Bonferroni corrected for multiple comparisons. Data from participants who scored 2.5 standard deviations above or below the mean accuracy were discarded. Data from two participants were excluded based on this criterion (both had overall accuracies lower than 40%, with some conditions having 0%). Data from an additional participant were excluded due to a program malfunction, resulting in 47 participants being included in the final analysis.
When necessary, F values were subjected to the Huynh-Feldt correction for the violation of the assumption of sphericity.

First, there was a main effect of presentation time [F(3,138)= 753.24, p<.001, $\eta^2_p = .94$], with accuracy increasing as presentation time increased. There was a significant main effect of level of schematization [F(4,184)= 241.68, p<.001, $\eta^2_p = .84$], indicating overall differences between levels of schematization. Follow-up comparisons revealed that each stimulus type was significantly different from all the others (all ps < .001), with highest accuracy for cartoons, followed by mid-cartoon faces, rotoscoped faces, mid-rotsocoped faces, and photos. This pattern of results supported my prediction that images become easier to process as they become schematized.

Crucially, there was an interaction between level of schematization and presentation time [F(12, 552)= 29.49, p<.001, $\eta^2_p = .39$]. To further probe this interaction, each stimulus set was compared in a separate ANOVA at each presentation time, with significance levels Bonferroni adjusted for multiple comparisons. For brevity, only significant differences are reported.

At 17 ms, accuracy for cartoon stimuli was higher than for all other stimulus categories. Accuracy for mid-cartoon and rotoscoped images was lower than for cartoon images, but higher than for photo and mid-rotsocoped images (ps<.001). At presentation times 33 ms and 50 ms, all stimulus sets differed from one another (ps<.01), with highest accuracy for cartoon images and lowest for photographic images. Finally, at the presentation time of 66 ms, of the three lowest accuracy stimulus sets (photo, mid-rotsocoped, and rotsocoped images), only accuracy for rotsocoped images and photos differed from each other.

Thus, the interaction revealed both a sharper increase in accuracy for the mid-rotsocoped and mid-cartoon stimulus sets between presentation time 17 ms and 33 ms, and a general
leveling out of accuracies at 66 ms. However, it is important to note that the order of conditions from highest accuracy to lowest was the same at all presentation times: cartoon, mid-cartoon, rotoscoped, mid-rotoscoped, and photo (see Figure 9).
The pattern of accuracy differences between stimulus types, particularly at the shortest presentation times, suggested the possibility that two low-level features, contrast and complexity, were contributing to accuracy of expression identification in ways which could easily be dissociated.

In a follow-up analysis, I collapsed across emotion and presentation time in all stimulus types except photographs. Within the remaining stimulus categories, the mid-rotoscope stimulus is a low-contrast version of the rotoscope stimulus, and the mid-cartoon a low contrast version of
the cartoon stimulus. That is, the contrast ratio (calculated as the relative RGB luminance of the lightest colour + .05 / the relative luminance of the darkest colour +.05) between the face outline and features and the face ‘skin-tone’ in the rotoscope and cartoon stimuli is always 21:1, or the maximum contrast on a computer monitor (i.e., pure black on pure white); the average contrast ratios for the mid-rotoscope and mid-cartoon stimuli are 7.5:1 and 12.8:1 respectively.

Likewise, the two cartoon stimulus types (cartoon and mid-cartoon stimuli) can be seen as lower featurally complex versions of the non-cartoon stimulus types (see Figure 10). To confirm this, I tested a separate sample of 20 participants on how complex or simple each of the image sets were, using a Likert scale from 1 (least featurally complex) to 7 (most featurally complex). The results of this manipulation check confirmed that the mid-cartoon (M= 1.91, SD= 1.38) and cartoon (M= 1.73, SD= 1.24) sets were in fact perceived as less complex than the mid-rotsocoped (M= 4.50, SD= 1.54) and rotoscooped sets (M= 3.77, SD= 1.59).

I thus grouped the mid-rotsocoped and mid-cartoon stimulus sets as “low contrast”, with the rotoscooped and cartoon as “high contrast.” Similarly, cartoon and mid-cartoon images were grouped as “low featural complexity” and rotoscooped and mid-rotsocoped images as “high featural complexity”. In this way, I could compare the main effect of contrast and the main effect of featural complexity independently across the stimulus sets, using a 2x2 repeated measures ANOVA (see Figure 10).

This analysis found a main effect of contrast [F(1,46)= 116.69, p<.001, η²_p = .72], as well as a main effect of featural complexity [F(1,46)= 408.77, p<.001, η²_p = .90], but no interaction between them (p>.250). This suggests that both the contrast and the complexity of an image affect accuracy (with higher contrast and lower complexity being represented as higher accuracies), but they do so without interacting with one another.
To examine potential effects of presentation time on the results, I next performed an ANOVA that included presentation time as an additional factor (3 within-subject factors: contrast, featural complexity, and presentation time).

Here there were interactions between presentation time and contrast \([F(3,144)= 61.93, p<.001, \eta^2_p = .56]\), and presentation time and complexity \([F(2.26,106.26)= 23.04, p<.001, \eta^2_p = .32]\), as well as an interaction between all three factors \([F(2.11,101.49)= 5.78, p<.001, \eta^2_p = .11]\).

Follow up comparisons for the presentation time by contrast interaction revealed that the interaction was driven by an attenuation of differences within a category type at longer stimulus.
presentation times. Critically, all comparisons comparing high contrast and low contrast images were significant at every presentation time (ps<.005). The interaction was instead driven by the comparison of the high contrast images at 50 ms and 66 ms presentation times for (p=.267). That is, there was no difference in accuracy between these two presentation times, but only for high-contrast images.

For the presentation time by complexity interaction, a similar pattern was found, where all comparisons of complexity at each presentation time were significant (ps<.001), so that high complexity images were always different from low complexity images at every presentation time. Instead, the interaction was driven by the comparisons of the longest two time windows (50 and 66 ms) for the low complexity images. In sum, these interactions show that features which promote discrimination of images, high contrast or low complexity, show less of a benefit at longer presentation times, and so may have hit a ‘ceiling’ of their usefulness between 50 and 66 ms presentation times. However, the influence of contrast and featural complexity were always significant at each presentation time.

This levelling-off effect within a stimulus set also characterized the three-way interaction. Comparisons between levels of contrast or levels of complexity were always significant at every presentation time (ps<.001). As with the previous interactions, this interaction was driven by a comparison of presentation times within stimulus sets. If the image was either high contrast or low complexity or both, the comparison of 50-66 ms time windows was non-significant. In the stimulus set with high contrast and low complexity, the comparison between the 33 ms and 50 ms presentation times was also non-significant. Essentially, this reflected a combination of what was found in the separate two-way interactions: for features that promote discrimination of images (high-contrast or low featural complexity), there is less of a benefit at longer presentation
times. However, just as in the smaller interactions, the influence of contrast and featural complexity were always significant at each presentation time.

Discussion

These results confirmed my prediction that images become easier to process as they become schematized – specifically as they become less complex and higher in contrast. This experiment found that this difference became more exaggerated at the shortest presentation times, showing for the first time a relative ease of expression discrimination for any type of face at 17ms. These results would suggest that cartoony faces differ from more realistic faces primarily in that their low-level features make them much easier to process, at least for expression information.

Given that both complexity and contrast influenced emotion discrimination in a dissociable manner, and that contrast has been shown previously to decrease ERP latencies (Y. Lu et al., 2014), in the next study I focused on complexity, using ERPs to further examine the influence of feature simplification on facial emotion processing.

Experiment 4

Rationale and prediction

Experiment 3 found that it is easier to discriminate emotional expressions on cartoony faces than photographs, and that this difference increases the more cartoony a face becomes. Building on these findings I next asked if the neural processes that are linked to better discriminability of cartoony faces are those associated with face-selective processing or low-level feature processing. Specifically, is the pattern of results observed in Experiment 3 indexed by an ERP that is sensitive to faces (the N170) or an earlier occipitally generated ERP (the P1)? The former outcome would suggest that cartoony faces may be less or more ‘face-like’ compared to photorealistic faces, and the latter result would suggest that cartoony faces differ from
photographic faces primarily due to their low-level feature differences. Note that ERP measurement demands that the presentation times in Experiment 4 be slower than in Experiment 3, which may reduce the difference between cartoony and photorealistic faces. Thus the present study represents a conservative test of the above question.

Participants

29 participants were recruited from the UBC psychology human subject pool (22 identified as female, mean age =19.9, 7 as male, mean age = 20.1). Here, again based on past studies performed by our research group, I considered a sample size of 25 as sufficient for finding meaningful ERP results. Four extra participants were included because they had already signed up to participate in redundant timeslots on the final week of testing (i.e., I concluded data analysis at the end of the week of the 25th participant). Two participants were excluded due to excessive EEG artifacts or poor performance. As with the previous experiment, participants were given course credit in exchange for participation.

Procedure

All stimuli were identical to those in experiment one, except that the mid-cartoon and mid-rotosced stimulus sets were discarded for two reasons: To maximize the number of trials recorded in ERP version of the task and to focus on the influence of contrast and complexity in facial emotion discrimination.

This left three stimulus sets: photo, rotosced, and cartoon faces. Whereas the comparison of cartoon with rotosced and photo stimulus sets would reveal the impact of featural complexity on face processing, the comparison of rotosced and photo images would demonstrate effects of contrast.

All images were also presented in blocks, so that between breaks only one stimulus set was shown at a time. In addition, I eliminated the use of masks to allow for a clean ERP
response, and all stimuli were presented for 500 ms to ensure that a complete waveform would be detected. That is, so that I would have the maximum amount of trials to average into ERPs, I wanted participants to be at ceiling as well as for the ERPs to my stimuli to be clean (and so, successful trials).

To further allow sufficient trials for ERP averaging, I also presented more trials for each stimulus type than in Experiment 3. Here 360 trials of each stimulus type were presented, in randomized blocks of 120 trials each, over the course of an hour for a total of 1080 trials.

The task participants were given was identical to the behavioural task. Participants responded to which emotion (Happy, Shocked, Disgust, Neutral) was expressed by each face after it was presented. However, to prevent response noise in the EEG data, participants were told not to respond until they reached the response screen, which occurred immediately after the 500 ms presentation time of the stimulus, although they were told to answer as quickly and accurately as possible once the response screen was shown. Due to the 500 ms presentation time compared to the 17-66 ms presentation times of the behavioral task, this task was designed to be much easier.

**EEG data acquisition**

Scalp recorded EEG data was recorded using a 64 channel Biosemi Actiview system. All EEG data was processed using the Matlab toolbox, ERPLAB (Lopez-Calderon & Luck, 2014). Continuous EEG data was recorded at a sampling rate of 512 Hz, and was band-pass filtered using an IIR Butterworth filter with half-amplitude cut-offs at 30 and 0.1 Hz during offline preprocessing. All data were referenced to two additional electrodes, which were placed on the left and right mastoid processes. In addition, two electrodes were placed on the outer canthi of both eyes, and one electrode was placed below the right eye. These three electrodes were used to
record eye movement and blink information so that trials containing eye movements could be discarded.

Epochs for sectioning continuous data into ERP bins were 500 ms in length, time-locked to stimulus onset, and were referenced to 200 ms pre-stimulus onset. Data was processed automatically for artifacts using a moving window peak-to-peak method with the epoch length as the test period and window of 100 ms, resulting in less than 10% of total trials being rejected. Finally, data from correct responses for each type of trial (e.g., disgust cartoon, disgust photo….) were averaged together for each participant.

Peak P1 activation was extracted from each averaged epoch by measuring local peak amplitude at electrode Oz and local peak latency within a window of 80-160 ms after stimulus onset. If no peak was reliably found using this method, that participant was excluded. This resulted in two participants being excluded from P1 analyses, leaving 25 participants.

For a P1-N170 peak-to-peak analysis described below, the P1 and N170 were extracted using local peak amplitude and local peak latency at electrodes P9 and P10 within a window of 150-220 ms. For this analysis two participants who did not have P1s or N170s that could be reliably extracted were excluded, leaving a total of 25 participants.

**EEG Analysis**

*P1* The P1 is observed at posterior sites, and is typically maximal at electrodes contralateral to where an attended stimulus was presented in the visual field (e.g., Luck, Heinze, Mangun, & Hillyard, 1990). As the stimuli were presented directly in the center of the visual field, I performed the initial P1 analyses on electrode Oz, which was also the electrode that showed the highest P1 amplitude in a grand average across all participants and conditions.
Peak to peak analysis. A further question concerned whether differences between conditions reflected processes specific to face processing (i.e., represented by differences in the N170 independently from the P1) or simply low-level featural changes in the stimulus sets carried over into the N170. For this secondary analysis, I next performed a peak-to-peak comparison to determine if the earlier P1 component was contributing to the N170 at the same electrode sites. Here P1 and N170 peaks were extracted at sites where the N170 was maximal. The N170 ERP component is typically strongest at 6 occipito-temporal electrodes (P7, P9, P07, and P8, P10, P08) (Sagiv & Bentin, 2001). However, only the most ventral two of these electrodes allowed reliable extraction of N170 and P1 peaks: electrodes P9 and P10. Although all six electrodes listed above are common targets for N170 peak extraction, it is also common to find the clearest or strongest N170 at P9/P10 (e.g., Fisher, Towler, & Eimer, 2015; Itier, Van Roon, & Alain, 2011). Thus, the peak-to-peak analyses were performed on data extracted only from these two electrode sites.

ERP epochs were averaged for each stimulus type (cartoon, rotoscoped, photo), for each expression (disgusted, happy, shocked, and neutral) and for the three electrode sites (Oz, P9 and P10).

Results

Behavioural Data

3x4 (stimulus type by emotional expression) repeated measures ANOVAs were performed on both accuracy and reaction time data. For accuracy, there was a main effect of stimulus type [F(1.4, 39.14)= 40.63, p<.001, η²p = .59]. Follow up comparisons revealed that greater accuracy for cartoon faces (98.2%) than for rotoscoped (93.4%)(p<.001) or photo (94.0%)(p<.001) images. The difference between rotoscoped and photo images, as a measure of
contrast, was not significant (p=.28), suggesting higher contrast on its own did not facilitate performance.

For reaction time, there was a main effect of stimulus type [F(2, 56)= 15.75, p<.001, η²_p = .36]. Planned comparisons revealed faster responses to the cartoon images (Mean RT = 767 ms) than to the rotoscoped (989 ms) (p<.001) and photo images (956 ms) (p=.001), mirroring the accuracy results. There was no difference between the rotoscoped and photo images (p=1.0) again indicating contrast did not have a singular effect.

These behavioral data show a clear pattern that the cartoon images were overall easier to respond to, even when all the images were presented at 500 ms.
**Figure 11. N170 data at sensors P9 and P10.** The waveforms on the left show the three stimulus sets averaged across all participants. The charts on the right represent the latencies and magnitude of those individual peaks averaged together; for latency, the Photo baseline stimulus N170 was slower than the Cartoon and Rotoscoped sets; for magnitude, the cartoon stimulus set N170 was greater than the Rotoscoped and Photo sets.

**ERP data**

For all analyses, pairwise contrasts were Bonferroni adjusted for multiple comparisons, and main effects and interactions were corrected using the Huynh-Feldt correction for the violation of sphericity where necessary. For brevity, only significant effects are reported;
however, all values for each condition for both P1 and N170 can be found at the end of the chapter. The data discussed here is illustrated in Figure 11.

**P1**

A 3x4 (stimulus type x expression) repeated measures ANOVA was performed on P1 amplitude and latency at electrode Oz. Analysis of P1 amplitude showed a main effect of stimulus type \[F(2,50)= 6.13, p=.004, \eta^2_p = .20\]. Follow up comparisons revealed that amplitudes elicited by cartoon images, which were highest in contrast and lowest in complexity, were significantly lower than those elicited by photographic images which were highest in both (p=.021). Comparison of amplitudes elicited by photo and rotoscoped images, as a measure of contrast, revealed only a trend level difference (p=.07). Comparison of cartoon and rotoscoped images, as a measure of complexity, revealed no difference (p=.829). These results suggest that, whereas combined reductions of contrast and complexity in cartoon images result in a smaller P1, neither is a singular driving factor in cortical facilitation by iconic images.

Analysis of P1 latency revealed no significant results (ps >.24).

**Effects of expression.**

Although I included emotional expression as a factor in the analysis, in accordance with experiment 1, analysis of expression simply served to illustrate that, for amplitude, the reported effects were slightly stronger for disgusted faces, with no significant effects for latency (ps>.05) (see the tables at the end of the chapter).

**Peak to peak analysis**

To help understand whether the pattern of P1 results reported above was carried forward into the face-sensitive N170, I performed a follow-up analysis in which I compared the amplitude and latency of the P1 peak with the N170 peak to see if these effects observed in the
N170 originated at the P1. A 3x4x2 (stimulus type x expression x electrode site) repeated measures ANOVA was performed on the peak-to-peak difference between the N170 and the P1 components.

Amplitude. The effect of stimulus type was not significant (p=.112), although there were some effects of emotional expression. This suggests that the pattern of results reflects a P1 effect that persists and influences the amplitude patterns observed for the N170. As such, these effects likely do not reflect processes specific to face recognition, but simply the low level featural differences between the stimuli.

Latency. None of the main effects or interactions were significant (all ps>.1), further confirming that effects of stimulus-type on the N170 were not face-specific but had an onset at the earlier P1.

Discussion
This study built upon Experiment 3 by showing neural correlates of the behavioural results I found in the earlier experiment – although it is worth noting that, as the trials were presented for longer times, this correlation presupposes that the different presentation times between experiments are not impacting the pattern of results between my earlier behavioural experiment and this ERP experiment.

The ERP results I found were confined to P1 results, and the different image types used did not impact the N170. What this means is that, contrary to some past studies showing N170 differences between schematic and photorealistic faces (Churches et al., 2014), the differences I found between the different types of faces were likely all driven by low-level features. This would suggest that cartoony faces are not less or more ‘face-like’ than photographic faces, but that low level features such as contrast and complexity contribute to them being processed more easily.
The difference between P1 effects and N170 effects can be difficult to tease apart. If the P1 prior to the N170 is influenced by the experimental manipulation, and the same pattern of results is found for the N170, it is possible that the N170 effect is simply the P1 effect carried forward (and this is typically how matching P1 and N170 effects would be considered). In this case, it may be that low-level features passed downstream from the P1 would also result in altered face-sensitive processing, making the two not entirely separable (Rossion & Caharel, 2011). However, a stimulus that elicited such an occurrence would not be considered as creating a true ‘N170 effect,’ at least in terms of face-sensitive processing. For this reason, it has become common for researchers to try and separate P1/N170 effects in the manner done here, by subtracting the N170 values from P1 values to get a difference score, which reveals whether magnitude changes are specific to the N170 or something that is being carried from the earlier P1 (e.g., Rossion & Caharel, 2011; Naumann, Senftleben, Santhosh, McPartland, & Webb, 2018).

General Discussion

The present study examined the hypothesis that iconic faces are processed differently from realistic depictions, and communicate emotion more effectively. In Experiment 3, I found that greater schematization in my stimuli, accomplished through increasing contrast and decreasing featural complexity, increased the likelihood a participant would be able to discriminate which emotional expression was present. In addition, both factors increased accuracy in a way that represented separable effects. Notably, at the fastest presentation times, discrimination of expressions in the simplest images had an accuracy rate nearly 60% higher than in realistic faces, and accuracy was substantially higher than is typically reported at such presentation times (Pessoa et al., 2005; Whalen et al., 1998).
Experiment 4 revealed that cartoon stimuli evoked distinct patterns of cortical processing from photographic stimuli. Thus, stimuli characterized by both higher contrast and lower levels of featural complexity were associated with lower P1 amplitudes; however, the lack of significant differences between photo and rotoscoped images as a measure of contrast, and cartoon and rotoscoped images as a measure of complexity, suggest neither feature singly accounted for the effect. Together these two experiments indicate that the greater contrast and simplicity of cartoon images facilitate rapid discrimination of facial emotion associated with reduced need for cortical processing of the images.

The P1 is known to be sensitive to low-level visual features of faces and objects, and is thought to represent an early step in a feed-forward volley of visual stimulus processing, in which information is extracted and passed on to higher-level regions of the ventral visual stream (Desjardins & Segalowitz, 2013; Kappenman & Luck, 2012; Rossion & Caharel, 2011; Woodman, 2010). These data are consistent with previous findings on pattern-evoked ERPs where decreased complexity was associated with decreased P1 magnitude (Oken et al., 1987). Recent research has also found that the P1 is magnified for faces above objects even when these stimuli are presented too quickly to be reportable (Mitsudo, Kamio, Goto, Nakashima, & Tobimatsu, 2011). This finding provides further convergent evidence that the advantage in discriminating information from more iconic images, based on low-level features, is heightened when images are presented at the threshold of discrimination and would first be observed at the latency of the P1. Such an advantage may then contribute to later configural processes that may modulate the face-sensitive N170 (Rossion & Caharel, 2011). This may help describe how previous work has found more negative N170 amplitudes for emoticons than photographs.
(Churches et al., 2014); while I replicate this effect with cartoon stimuli, the findings indicate that this N170 modulation is not N170 specific, but has an onset much earlier at the P1.

Together with the above previous studies, the results of my experiments tell a story: increased contrast and complexity enhance the P1, which is a measure of relatively early perceptual processing of stimulus features. This enhancement can be seen as propagating forward into later stages of face-sensitive processing, and are observed in the N170 peak as an ongoing reflection of featural differences. This is then reflected in the behavioral results as decreased accuracy and (in Experiment 4) longer reaction times. The finding that differences between conditions observed in the N170 originated with the P1 is informed by a longstanding discussion about the role of the P1 in face processing.

P1 effects that precede and correspond to later N170 effects have been frequently observed – for example, contrast reversal and inversion have been found to affect P1 as well as N170 latency and magnitude (Itier & Taylor, 2004; Margot J. Taylor, 2002). The caveat to such findings is that inverting the contrast or spatial orientation of a face image, as with any other image, is directly influencing the stimulus’ low level features in a manner similar to this study. The simplest explanation is that the changes in P1 amplitude found here and in previous studies represents a difference in availability of information in a given stimulus, which can facilitate subsequent processing. My findings suggest that exaggeration of low-level features provides the basis for more efficient extraction of facial information in general, including information about facial identity (e.g., caricatures). It is also possible that the effects of schematization in facilitating facial emotion identification may not be specific to face processing but may generalize to ease of information extracted from any type of stimulus with higher contrast and simpler features.
It is important to note, however, that in creating images with low complexity, I also reduced stimulus variability and such reduced variability may also have influenced my results. Despite efforts to add variability to the cartoon stimuli, it can be difficult to mirror the natural variability in photographic faces in the medium of cartoons. Future research can investigate questions of stimulus variability as a contributing factor to the effects of iconic stimuli. From the results found in this study, however, arises the question of whether ‘schematization’ can also enhance the efficiency of object discrimination, or whether it bestows a special advantage to configural processes that are particularly important for discrimination of facial identity and emotion.

In practice, contrast and featural complexity are frequently manipulated together, such as in public environments when information must be detected quickly across contexts (e.g., Babbitt Kline, Ghali, Kline, & Brown, 1990). However, graphic emoticons have decreased in contrast since their inception (e.g., colour emoticons have lower contrast than their ASCII equivalents) but remain featurally simplistic, underscoring the importance of this factor. Aside from simplicity, emoticon features vary widely, and are still widely used for modulating textual emotional content (Rojas, Kirschenmann, & Wolpers, 2012; Luor, Wu, Lu, & Tao, 2010). The communicative role of other features, such as colour, in iconic representations, is another interesting avenue for research. Yet other questions to be pursued include, but are not limited to, whether other factors (e.g., emotional intensity) were manipulated secondarily along with contrast and featural complexity, and if so, what, if any role do they play in the present results. Similarly, in this study I tested low-level features that varied qualitatively, and so there could be effects of contrast or featural complexity that could only be found by scaling both low-level features parametrically. Further research will be required to investigate this possibility.
Iconic faces can be viewed either as analogous to realistic images or as a distinct class of stimulus. The findings in this chapter support the view that iconic representations serve a distinct role - to impart specific information quickly and efficiently - and highlight the advantages of simplifying image features and increasing contrast to communicate emotion. In addition, my data suggest that the effects of iconization may not be specific to faces, but rather to any stimulus that has these low-level featural changes. It is thus important to consider that such features are not just potential low-level confounds but contribute to specific communicative functions. However, it is unknown if the discrimination of more subtle real-world types of emotional expression would also benefit from iconic representation (e.g., the ‘Duchenne’ smile, where genuine happiness is expressed with the wrinkling of the corners of the eyes) (Ekman, Davidson, & Friesen, 1990). It may be that iconic images have a communicative advantage only for simple visual information, a hypothesis that invites future research.
Effects of emotional expression and excluded results
Experiment 3

There was a main effect of expression type $[F(3,138)= 12.42, p<.001, \eta^2 p = .21]$, suggesting that some emotions may be easier to process than others. However, follow-up comparisons revealed that this was driven purely by disgust, which was significantly different from all the other emotions (all $p$s<.001, with the exception of the comparison with neutral expression, $p=.032$). No other expressions differed from each other (all $p$s>.5). This likely simply indicates that the pattern of results was magnified for disgust compared to the other emotions.

There was also an interaction between level of schematization and expression $[F(12,552)= 6.27, p<.001, \eta^2 p = .12]$. Planned contrasts revealed this interaction was driven by faster identification of disgust faces at shorter presentation times.

Finally there was a three-way interaction between stimulus type, expression, and presentation time $[F(36,1656)= 11.55, p<.001, \eta^2 p = .20]$. This interaction was driven once again by an exaggerated profile for disgust compared to the other expressions (the difference in accuracy between the disgust and other stimulus categories was wider at the faster presentation times).

Experiment 4 Behavioral

For accuracy, there was a main effect of expression type $[F(2.4, 66.64)= 9.08, p<.001, \eta^2 p = .25]$. While the disgust expression was different from happy ($p=.04$), this effect was primarily driven by the shocked expression differing from both happy ($p<.001$) and neutral ($p=.001$). No other comparisons were significant ($p$s>.250). The shocked faces may have seemed more ambiguous, as at 92.9%, they were over 3% lower than happy (97.4%) and neutral (96.5%) faces.
There was also an interaction between stimulus type and expression type [F(4.39, 122.90) = 6.74, p < .001, η²p = .194]. Follow up contrasts revealed that this interaction was driven by the relative flatness of cartoon images (all ps = 1.0, with a range of 98.0 – 98.5%) compared to the other stimulus categories. For the rotoscoped images, the shocked expressions were significantly lower (90.0%) than the happy (96.8%) and neutral images (95.2%) (ps = .001 and .01 respectively). Similarly, for photos, the shocked expressions (90.1%) were significantly lower than the happy (97.2%) and neutral (96.2%) images (ps < .001). However here the disgust faces (92.4%) were also low enough compared to the happy faces to be significant (p = .048).

Altogether this pattern of accuracy to emotional expressions suggested that the shocked and disgusted faces were easier to mistake than the happy or neutral, or that, more likely, with more realistic faces there is simply a bias to responding happy or neutral versus shocked or disgusted.

For reaction times, there was also a main effect of expression type [F(3, 84) = 4.41, p = .006, η²p = .14]. Planned comparisons revealed that this was driven by happy images being responded to faster than disgust (p = .024) or shocked (p = .012) images.

There was no interaction between stimulus type and expression type (p = .475).

P1 - Oz

Expressions

For P1 amplitude at electrode Oz, there was also a main effect of expression [F(3, 75) = 4.83, p < .003, η²p = .16]. Follow up comparisons revealed that this was driven by the unique pattern of disgust faces, which were significantly different from shocked faces (p = .02) and neutral faces (p = .01). No other comparisons were significant (ps > .37). There was also a significant interaction of expression type and stimulus type [F(4.66, 116.48) = 2.58, p = .033, η²p = .09].
There were no significant expression effects on the P1 for latency at Oz, or for amplitude or latency at P9/P10 (ps>.05).

**P1 – P9/P10**

Analysis of P1 amplitude revealed a main effect of stimulus type \[F(1.56,42.16)= 7.83, \ p=.003, \ \eta^2_p = .23\]. Follow up comparisons revealed that cartoon images evoked a lower amplitude P1 than photos (p=.049) and rotoscoped (p<.001) images, but that photo and rotoscoped images did not significantly differ.

There was also a main effect for the sensor site at which P1 was recorded \[F(1, 27)= 21.60, \ p<.001, \ \eta^2_p = .44\], with P10 showing a larger P1 than P9.

Analysis of P1 latency revealed a main effect of stimulus type \[F(1.44, 38.87) = 48.03, \ p<.001, \ \eta^2_p = .64\]. Pairwise comparisons revealed that latencies for each of the three stimulus types different from each other (all ps <.01). Cartoon images evoked the shortest latencies, followed by rotoscoped images, and photos evoked the longest latencies. No other main effects or interactions were significant (all ps>.1)

**N170 – P9/P10**

*Amplitude*

Analysis of N170 amplitude revealed a main effect of stimulus type \[F(1.63,42.29)= 8.88, \ p=.001, \ \eta^2_p = .25\] (see Figure 5). Contrasts revealed that the N170 for the cartoon stimulus set was larger than both the rotoscoped and photo sets (p=.017 and p=.004, respectively), but that these latter two sets did not differ from each other (p=.167). Because the rotoscoped images and the cartoon images share contrast, this finding suggests that featural simplicity of the cartoon stimulus set rather than differences in contrast underlie differences in N170 amplitude, which matches the differences found in the P1 analyses.
There was also an interaction between electrode and stimulus set \([F(2, 52)= 14.05, p=.03, \eta^2_p = .13]\); Follow-up ANOVAs at each electrode site (i.e., one for P9 and one for P10) revealed that this is because, while the main effect of stimulus set was in the expected direction, it was not significant at P9 when that electrode was considered alone \((p=.11)\); in contrast, the effect was robust at P10 \([F(2, 56)= 11.19, p<.001, \eta^2_p = .29]\).

**Amplitude (expressions)**

There was also a main effect of expression type \([F(3,78)=12.58, p<.001, \eta^2_p = .33]\). Pairwise comparisons revealed that the disgust images evoked a larger amplitude N170, which differed from that for all other expressions \((p>.05)\), which in turn did not differ from each other \((all \ p > .2)\).

There was an interaction between stimulus set and expression type \([F(6, 156)= 5.07, p=.009, \eta^2_p = .10]\). Planned contrasts revealed that the main effect of stimulus set was not significant for the neutral expression stimuli \((p=.143)\) but was significant for all other expressions \((all \ Fs > 4.0, all \ p< .05)\), suggesting that the advantage of featural complexity may have a greater advantage in emotional (i.e., non-neutral) stimuli.

**Latency**

Analysis of N170 latency revealed a main effect of stimulus type \([F(2, 52) = 59.46, p<.001, \eta^2_p = .70]\). Contrasts revealed differences in latency between the N170 for photos and both rotoscoped and cartoon stimuli \((ps <.001)\). The rotoscoped and cartoon stimuli evoked an N170 that was, on average, 13.9 ms faster than the photo stimuli.

**Latency (expressions)**

We observed a main effect of expression type \([F(3, 78)= 2.981, p=.036, \eta^2_p = .10]\). Pairwise contrasts revealed that, although happy expressions were processed slightly faster than
other expressions (on average 1.56 - 2.66ms), there were no significant latency differences between pairs of expression type (ps>.05).

Peak to Peak Expressions

A main effect of expression was found \[F(2.25, 62.03)= 19.49, \ p<.001, \ \eta^2_p =.44\]. Pairwise comparisons revealed that this was driven by the disgust faces showing a larger difference than all other stimulus types (ps<.001). This further supports our disgust stimuli as being processed uniquely, but as this did not vary with stimulus type, this information is tangential to our hypothesis.

There was also peak to peak interaction of stimulus type and emotional expression, \[F(6, 150)= 3.63, \ p=.002, \ \eta^2_p =.13\]. However, in follow up analyses this interaction seemed to be driven solely by a significantly greater N170/P1 difference for happy faces only between rotoscoped and photo stimulus sets (p=.044).
Table 2.
Full accuracies for Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Photo</th>
<th>Mid-rotoscopyed</th>
<th>Rotoscope</th>
<th>Mid-cartoon</th>
<th>Cartoon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presentation time (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>33</td>
<td>50</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>Mean</td>
<td>.173</td>
<td>.171</td>
<td>.234</td>
<td>.506</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.024</td>
<td>.028</td>
<td>.032</td>
<td>.041</td>
</tr>
<tr>
<td>Happy</td>
<td>Mean</td>
<td>.267</td>
<td>.481</td>
<td>.558</td>
<td>.730</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.032</td>
<td>.036</td>
<td>.032</td>
<td>.041</td>
</tr>
<tr>
<td>Shocked</td>
<td>Mean</td>
<td>.483</td>
<td>.685</td>
<td>.658</td>
<td>.861</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.041</td>
<td>.041</td>
<td>.039</td>
<td>.026</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean</td>
<td>.600</td>
<td>.827</td>
<td>.789</td>
<td>.886</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.034</td>
<td>.036</td>
<td>.024</td>
<td>.030</td>
</tr>
<tr>
<td>Disgust</td>
<td>Mean</td>
<td>.243</td>
<td>.234</td>
<td>.491</td>
<td>.577</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.032</td>
<td>.034</td>
<td>.032</td>
<td>.042</td>
</tr>
<tr>
<td>Happy</td>
<td>Mean</td>
<td>.437</td>
<td>.558</td>
<td>.755</td>
<td>.730</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.036</td>
<td>.026</td>
<td>.035</td>
<td>.034</td>
</tr>
<tr>
<td>Shocked</td>
<td>Mean</td>
<td>.559</td>
<td>.786</td>
<td>.729</td>
<td>.807</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.023</td>
<td>.020</td>
<td>.030</td>
<td>.025</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean</td>
<td>.612</td>
<td>.882</td>
<td>.779</td>
<td>.803</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.019</td>
<td>.019</td>
<td>.030</td>
<td>.029</td>
</tr>
<tr>
<td>Disgust</td>
<td>Mean</td>
<td>.506</td>
<td>.547</td>
<td>.491</td>
<td>.577</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.041</td>
<td>.034</td>
<td>.032</td>
<td>.042</td>
</tr>
<tr>
<td>Happy</td>
<td>Mean</td>
<td>.588</td>
<td>.730</td>
<td>.755</td>
<td>.777</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.032</td>
<td>.026</td>
<td>.035</td>
<td>.035</td>
</tr>
<tr>
<td>Shocked</td>
<td>Mean</td>
<td>.678</td>
<td>.861</td>
<td>.829</td>
<td>.807</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.020</td>
<td>.020</td>
<td>.030</td>
<td>.025</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean</td>
<td>.655</td>
<td>.886</td>
<td>.779</td>
<td>.803</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.029</td>
<td>.029</td>
<td>.030</td>
<td>.029</td>
</tr>
<tr>
<td>Disgust</td>
<td>Mean</td>
<td>.903</td>
<td>.543</td>
<td>.620</td>
<td>.577</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.019</td>
<td>.040</td>
<td>.036</td>
<td>.042</td>
</tr>
<tr>
<td>Happy</td>
<td>Mean</td>
<td>.948</td>
<td>.786</td>
<td>.846</td>
<td>.777</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.011</td>
<td>.034</td>
<td>.027</td>
<td>.035</td>
</tr>
<tr>
<td>Shocked</td>
<td>Mean</td>
<td>.961</td>
<td>.854</td>
<td>.879</td>
<td>.807</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.011</td>
<td>.033</td>
<td>.023</td>
<td>.021</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean</td>
<td>.965</td>
<td>.863</td>
<td>.881</td>
<td>.803</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.009</td>
<td>.032</td>
<td>.020</td>
<td>.013</td>
</tr>
<tr>
<td>Disgust</td>
<td>Mean</td>
<td>.739</td>
<td>.870</td>
<td>.581</td>
<td>.903</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.035</td>
<td>.022</td>
<td>.043</td>
<td>.019</td>
</tr>
<tr>
<td>Happy</td>
<td>Mean</td>
<td>.886</td>
<td>.922</td>
<td>.871</td>
<td>.948</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.016</td>
<td>.016</td>
<td>.021</td>
<td>.011</td>
</tr>
<tr>
<td>Shocked</td>
<td>Mean</td>
<td>.939</td>
<td>.939</td>
<td>.912</td>
<td>.961</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.014</td>
<td>.017</td>
<td>.017</td>
<td>.011</td>
</tr>
<tr>
<td>Neutral</td>
<td>Mean</td>
<td>.960</td>
<td>.946</td>
<td>.943</td>
<td>.965</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>.011</td>
<td>.014</td>
<td>.013</td>
<td>.009</td>
</tr>
</tbody>
</table>
Table 3.

ERPs full data (N170)

<table>
<thead>
<tr>
<th>Stimulus set</th>
<th>Expression</th>
<th>Sensor</th>
<th>Magnitude (μV)</th>
<th>Latency (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Std. Error</td>
</tr>
<tr>
<td><strong>Photo</strong></td>
<td>Disgust</td>
<td>P9</td>
<td>-1.024</td>
<td>.483</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-.357</td>
<td>.583</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>P9</td>
<td>-.756</td>
<td>.451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>.037</td>
<td>.615</td>
</tr>
<tr>
<td></td>
<td>Shocked</td>
<td>P9</td>
<td>-1.273</td>
<td>.467</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>.244</td>
<td>.606</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>P9</td>
<td>-.808</td>
<td>.508</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>.007</td>
<td>.586</td>
</tr>
<tr>
<td><strong>Rotoscoped</strong></td>
<td>Disgust</td>
<td>P9</td>
<td>-1.195</td>
<td>.417</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-1.504</td>
<td>.580</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>P9</td>
<td>-.877</td>
<td>.408</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-.757</td>
<td>.578</td>
</tr>
<tr>
<td></td>
<td>Shocked</td>
<td>P9</td>
<td>-.981</td>
<td>.394</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-.708</td>
<td>.596</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>P9</td>
<td>-1.041</td>
<td>.448</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-.738</td>
<td>.552</td>
</tr>
<tr>
<td><strong>Cartoon</strong></td>
<td>Disgust</td>
<td>P9</td>
<td>-2.192</td>
<td>.553</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-2.354</td>
<td>.757</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>P9</td>
<td>-1.186</td>
<td>.406</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-.947</td>
<td>.594</td>
</tr>
<tr>
<td></td>
<td>Shocked</td>
<td>P9</td>
<td>-1.726</td>
<td>.387</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-1.666</td>
<td>.680</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>P9</td>
<td>-.757</td>
<td>.450</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>-.751</td>
<td>.471</td>
</tr>
<tr>
<td>Stimulus Set</td>
<td>Expression</td>
<td>Sensor</td>
<td>Magnitude (μv) Mean</td>
<td>Magnitude (μv) Std. Error</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>--------</td>
<td>---------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Photo</td>
<td>Disgust</td>
<td>P9</td>
<td>3.132</td>
<td>1.844</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.061</td>
<td>2.503</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>P9</td>
<td>2.987</td>
<td>1.719</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.171</td>
<td>2.424</td>
</tr>
<tr>
<td></td>
<td>Shocked</td>
<td>P9</td>
<td>2.506</td>
<td>1.834</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.085</td>
<td>2.584</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>P9</td>
<td>2.787</td>
<td>1.613</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.112</td>
<td>2.282</td>
</tr>
<tr>
<td>Rotoscoped</td>
<td>Disgust</td>
<td>P9</td>
<td>3.043</td>
<td>1.521</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.179</td>
<td>2.713</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>P9</td>
<td>3.234</td>
<td>1.581</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.056</td>
<td>2.519</td>
</tr>
<tr>
<td></td>
<td>Shocked</td>
<td>P9</td>
<td>3.029</td>
<td>1.792</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>4.990</td>
<td>2.523</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>P9</td>
<td>2.881</td>
<td>1.846</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>5.179</td>
<td>3.093</td>
</tr>
<tr>
<td>Cartoon</td>
<td>Disgust</td>
<td>P9</td>
<td>2.623</td>
<td>1.637</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>4.035</td>
<td>2.776</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>P9</td>
<td>2.490</td>
<td>1.654</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>4.250</td>
<td>2.540</td>
</tr>
<tr>
<td></td>
<td>Shocked</td>
<td>P9</td>
<td>2.280</td>
<td>1.500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>4.203</td>
<td>3.053</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>P9</td>
<td>2.851</td>
<td>1.659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P10</td>
<td>4.535</td>
<td>2.707</td>
</tr>
</tbody>
</table>
Chapter 4: Cartoony and photorealistic upfix dyads

Introduction

In the popular children’s comic Asterix and Obelix, the reader is presented with a series of images of elements that resemble objects in the real world. There are figurative, if exaggerated, human characters, animals, and landscapes. However, this reliance on verisimilitude disappears when Asterix becomes angry. As he grows angry, black swirling lines appear around his reddened face. These lines, which represent his emotion, are derived from a “vocabulary” within a unique visual language used in comics (e.g., Cohn et al., 2016; Walker, 1980). While this visual language vocabulary originates in comics and animation, such signs also appear in other visual media, such as emoji. Curiously though, these forms generally appear only when media are non-realistic—we rarely see black swirling lines above a photographed face. This raises the following question: If the characters and landscapes in comics typically are meant to resemble those in the real world, why do they uniquely employ the more abstract symbolism of visual language?

Many such examples of conventionalized signs for representing emotions or ideas appear across different styles of comic and cartoon media. In the world of cartoon characters, hearts can circle a character’s head when they are in love, clouds when they are depressed, and birds when they are dizzy. Some of these images could have natural analogs in realistic media, such as those depicting clouds or birds, whereas others seem by nature to be unique to drawings, such as hearts or radiating black lines. Collectively, these symbols are common enough that taxonomies have been attempted to describe them, originally in humorous ways (Walker, 1980), but more recently embedded in the cognitive and linguistic sciences (Cohn, 2013a; Cohn & Ehly, 2016; Cohn et al., 2016a). Within these latter taxonomic efforts, these ‘above the head’ signs specifically have
been called ‘upfixes’, because they are visual ‘affixes’ that are ‘up’ from a head (Cohn, 2013a; Cohn et al., 2016a). (Note that affixes are present not only in written language, but verbal language as well. Upfixes, on the other hand, are usually discussed as a visual phenomenon.)

Upfix dyads contain different types of both iconic and symbolic imagery. Though images like clouds or birds are iconic on their own (resembling the objects they represent), when put above a head they take on a unique symbolic meaning, adding another level of reference. This distinction may be important in studying why such signs appear predominantly in the realm of comics and animation. That is, perhaps the inclusion of symbolic representations is uniquely supported by cartoony images rather than realistic ones? Indeed, there is evidence of conventional stylistic patterns even in the iconic aspects of drawn images, although this is not as obvious as in aspects of the images that seem more intuitively ‘out of place’ such as the symbolic upfixes (Cohn, 2013a; Wilson, 1988).

Recently, the use of these signs for visual media, especially comic books, has come under the lens of scientific inquiry. Researchers have proposed that drawings, especially those found in comics, use structures similar to those in languages (Cohn, 2013a), and neurocognitive research has suggested that sequential image processing, as in comics, shares a neural pattern to how we perceive sentence structure (Cohn, 2013a; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012). Different ‘visual languages’ employ visual vocabularies of conventionalized signs—both iconic and symbolic—in comics across the world (Cohn, 2013a; Cohn & Ehly, 2016), and elements like upfixes involve spatial constraints not unlike linguistic lexical items (Cohn, 2013a; Cohn & Maher, 2015; Cohn et al., 2016a; Forceville, 2011). For example, upfixes that are moved next to a head and/or are incongruent with the facial expression are harder to understand than those above a head or agreeing with the facial expression—even for novel,
unconventionalized upfixes (Cohn et al., 2016a). Furthermore, such visual vocabulary items may also invoke similar conceptual metaphors as linguistic utterances (Eerden, 2009; Forceville, 2005, 2011, 2016). These findings suggest that understanding the use of different types of signs originating in comics may be related to their organization within a ‘visual linguistic system’, rather than simply due to visual dissimilarities between cartoon and realistic media.

A contrary possibility is that the meaning of these signs can only be processed sufficiently easily in a cartoon form due to low-level visual properties. That is, it is simply easier and faster to process information in a cartoon sign, regardless of its ‘linguistic’ status. There is some evidence that cartoon imagery is processed differently from more realistic imagery, which may encourage the use of symbolic signs such as upfixes. For example, contrast and complexity have been shown to be important aspects of extracting information from faces (Gray et al., 2013; Yue et al., 2011), and cartoon faces are more likely to be processed featurally than holistically (Prazak & Burgund, 2014). Thus, symbolic imagery within a cartoon may allow one to extract meaning more rapidly and accurately and require lower levels of sustained attention.

With these two possibilities – that cartoon images are the building blocks for a visual linguistic system, or that cartoon images are simply easier to process due to low level features – I constructed studies that collectively could yield evidence that favoured one account rather than the other by using two types of data. One possibility builds from the idea that cartoon images underpin the use of visual linguistic schemas in a way that photographic images do not, or cannot. From this it follows that the content of two images embedded within a visual linguistic schema, which have meanings that inform one another, would be understood differently if the images were cartoony rather than photographic. For example, upfix/face dyads, where a symbol above a facial expression forms a new holistic schema meaning not found in either individual
component, has been studied as one of the base ‘units’ of visual linguistic systems. This is because the face and symbol are only meaningful together when placed in certain spatial configurations, and are often only interpretable when presented together (Cohn et al., 2016a) – much as affixes are used in written or verbal language. Thus, the upfix/face dyad itself can be seen as a visual linguistic schema. Using this base visual linguistic unit of the face/upfix dyad, I could measure how easily people understand upfix/face combined meaning when the stimuli are cartoony versus photographic. If cartoony images underlie successful interpretation of the upfix/face dyad visual linguistic unit, participants should find it easier to recognize how the upfix/face symbolism fits together if the stimuli are cartoony than if they are photographic. One way to measure this is to see how easily participants recognize – using accuracy and reaction times - that the meaning of an upfix and face had congruent meanings versus incongruent meanings.

The alternative possibility, that cartoony images are used in visual languages (such as with upfixes) simply due to their ease of processing, is not something I would expect to find reflected in congruency effects. Instead, if cartoony images are simply easier to process, participants should devote less overt attention to cartoony elements of an upfix/face dyad than to photographic elements. A straight-forward way of measuring such attentional changes would be through the use of eye tracking. Measuring gaze direction can provide a direct means of examining whether readers attend to symbols differently depending on their degree of visual realism (cartoony vs. photorealistic). Eye tracking studies have long established gaze as a reliable measure of overt attention: The length of time a viewer looks at a stimulus is an index of how much attention that item requires, and the timing of an eye movement to a stimulus is a measure of stimulus salience (Hoang Duc, Bays, & Husain, 2008; Kowler, 2011). Thus, eye
tracking allows us to examine how aspects of visual depiction influence allocation of attention to specific elements of an image when observers must decode the meaning of learned symbols such as emotionally expressive faces with upfixes.

In the present study, I examined how an image’s level of realism influenced discrimination of upfixes/face dyads, specifically whether or not the upfix symbols were seen as ‘congruent’ with the expressions found on the faces. I created images composed of dyads containing emotionally expressive faces and upfixes, and manipulated the pairings so that either the face, upfix, or both were realistic or cartoony. I had two hypotheses. First, that upfix/face dyads that were entirely cartoony would be easier for participants to understand, as represented by higher accuracy and faster response times, when congruent in meaning versus incongruent, and that this effect would not be present for photographic dyads. Second, I hypothesized that cartoony items would require less viewing time than realistic images overall, so that items containing cartoony images would elicit processing advantages. In Experiment 5, I approached this question by asking participants to determine if the dyads were congruent (i.e., facial expressions matched the upfix meaning) or incongruent, and extended this work with evidence from eye-tracking to see how gaze was affected by the realism of dyadic elements. In Experiment 6, I used dyads which were more difficult to rate for congruency on a smaller set of symbols in an attempt to magnify behavioural and eye-tracking effects from Experiment 5.

Experiment 5

Rationale and prediction

My previous experiments suggest that cartoony faces are easier to discriminate between or process than photographic faces because of their low-level features. The present experiment
investigates an issue derived from the view that cartoony images are commonly used in visual media with a linguistic structure because they are especially able to accommodate linguistic interpretations (Cohn, 2007). In the case of upfix/face dyads, the dyad itself is commonly seen as a unit of visual linguistic structure, as understanding how an upfix and face relate to one another relies on understanding how a spatially located symbol above a head alters the meaning of a facial expression, in a similar manner to how a prefix can alter a word in written language by being attached to its beginning (Cohn et al., 2016a). Thus, the upfix/face dyad itself can be, and has been, described as ‘linguistic.’ But what I was more interested in was whether this unit of visual linguistic systems would be understood differently if the images were cartoony rather than photographic.

In Experiment 5, I tested the proposal that cartoony images encourage a linguistic schema interpretation – that is, that upfix/face dyads are interpreted as more meaningful together when presented as cartoony images. I did this by having participants respond to a face and an upfix together and by manipulating the congruency between the emotional expression of the face expression and its upfix. (Note again that an upfix is a symbol above a face (e.g., a stormy cloud), that relates to and informs on the expression on the face).

My reasoning was as follows: if face-upfix linguistic interpretations are particularly easy for cartoony images, then a face-upfix dyad that is cartoony should yield faster reaction times, higher accuracies, and show stronger congruency effects compared to a photorealistic face-upfix dyad. In other words, I expected that cartoony face-upfix dyads would be seen as relating more to one another, and so would be easier to understand together when congruent and harder when incongruent. If I found such a congruency effect with cartoony images and a smaller effect for photographic upfix/face dyads, it would support the idea that cartoony images might be
especially able to support linguistic interpretations. This is because the upfix/face dyad is touted as relying on a linguistic-type schema, and so if cartoony media help them to be understood, that medium would be more appropriate for using upfix/face dyads in visual linguistic system (e.g., such as comics). More importantly for the purpose of this dissertation, a particularly strong cartoony congruency effect would suggest that cartoony images can be differentiated from photographic images not only by their low-level features, but also in their ability to act as a building block for linguistic structures (for example, as in comics). To more thoroughly examine this research question I also tested performance when one item (face or upfix) was cartoony and the other was photorealistic.

Additionally, to understand the way that participants took in the face-upfix information I monitored participants' eye movements, to measure overt visual attention. As mentioned in the introduction, this was to assess the additional idea that cartoony images are simply easier to process due to their low-level featural differences. Barring finding any congruency effects, this study could provide convergent evidence that cartoony stimuli are used because they are easier to process than photographic images, and that this would appear in my study as less overt attention (i.e., eye-movements) to cartoony stimuli than to photographic stimuli.

Participants
33 undergraduate students (Mean age = 20.8, 28 women, 5 men) with normal or corrected to normal vision participated in exchange for course credit. The experiment was approved by the University of British Columbia’s institutional ethics review board, and each participant gave written informed consent.

Stimuli.
Two image elements were displayed on-screen in each trial, forming a dyad. The bottom image was always a facial expression (happy, angry, sad, or neutral), and the top image was always an
upfix (light-bulb, skull and crossbones, gears, or cloud) (see Figure 12 and Figure 13). Face stimuli were each 135 x 175 pixels, and upfixes were fit into transparent image frames 175 x 175 pixels.
**Figure 12. All the upfix stimuli used in Experiment 5.**

**Figure 13. All the types of upfix/face dyads, manipulating both the cartooniness of the separate parts as well as their emotional congruency.**
Following the manipulation of Cohn, Murthy, and Foulsham (2016), dyads could be either congruent (i.e., where the meaning of the upfix matched the meaning of the facial expression, such as a smiling face with a light-bulb) or incongruent (e.g., a smiling face with a cloud upfix). For the central manipulation, to examine how an image’s level of realism influenced responses, I created a photographic version and a cartoon analog for each facial expression. In turn, each photorealistic upfix also had a cartoon analog. Cartoon upfixes were constructed by taking photographic stimuli and tracing cartoon versions directly over them in GIMP, an image manipulation program (Kimball & Mattis, 1996).

Procedure.

The experimental design was adapted from Cohn, Murthy, and Foulsham (2016). Participants completed the experiment in a dimly lit testing room, with trials displayed on a computer screen with a neutral grey background. Figure 14 breaks down a typical trial.

![Figure 14. A typical trial. Participants were shown each image until response.](image)
With every combination of trials (4 upfixes x 4 faces x 2 variations of upfix x 2 variations of face) there were 64 trials. Each possible combination was presented twice, resulting in 128 trials. Using every combination yielded 3 possible combinations of dyads which would be incongruent, and one possible combination that was congruent (e.g., a lightbulb matches one expression, and does not match three). All trials were presented in random order.

First, participants were given a set of instructions which told them to indicate whether or not the expression in each trial matched the meaning of the accompanying upfix, responding as quickly and as accurately as possible with a key-press. For each trial, a fixation cross appeared for 300 ms in the center of the screen, after which the dyad appeared until response (pressing m for match and z for mismatch). Trials would not time out, but would wait for a response. Dyads appeared in the center of the screen as well, so that the upfix was above where the fixation cross had been, and the face part of the dyad was below. Immediately after a key press, a new trial started. This allowed for a rapid succession of trials that followed the pace of each individual participant.

_Eye-tracking_

Eye-tracking data was collected using an SMI desktop mounted eye tracker with a sampling rate of 500 Hz. Calibration was performed on the right eye of each participant prior to the start of their experimental session. Stimuli were displayed on an LCD monitor using SMI proprietary Experiment Builder software.

Two separate AOIs were defined. One covered the area of the face, and was a circle that would encompass each face stimulus. The other AOI surrounded the upfix, ensuring that the AOI window encompassed each of the differently shaped upfixes. The face AOI was a circle with a radius of 283 pixels (that is, comprised of an area of 251116 pixels), and the upfix AOI was a
rectangle that comprised an area of 141625 pixels. Size and placement of these two AOIs were held constant across all trials.

In addition to accuracy and response times, three eye tracking variables were analyzed: dwell time, number of fixations, and entry time. Dwell time is the amount of time spent within an AOI overall, whereas number of fixations refers to how many separate times gaze halted within the AOI overall. These measures are not entirely independent, and both are measures of attention allocation, with more fixations or dwell time indicating greater deployment of attention (Hoang Duc et al., 2008). Entry time, in contrast, is the absolute time in a trial at which an AOI was first entered by the participant’s gaze. This measure represents an estimate of how early attention was shifted, as an earlier entry time means that, on average, gaze was oriented to that AOI earlier.

Results

Analysis of all eye-tracking data, with the exception of entry time data, was performed on a ratio between AOI values (e.g., upfix dwell time / face dwell time). The ratio was calculated to better represent the proportion of dwell time or fixations attributed to the AOIs considered together rather than as the total values in each AOI. Entry time numbers had a much higher range than dwell time and fixations and so a ratio measure was not meaningful. That is, there were some trials where participants did not look at one part of the dyad until very late in the trial, vastly exaggerating any possible ratio.

2 x 4 repeated measures ANOVAs were performed on dwell time and number of fixations, with congruency (congruent vs. incongruent trials) and dyad-type (cartoon face/cartoon upfix, cartoon face/photo upfix, photo face/cartoon upfix, photo face/photo upfix) as within-subject factors. The same factors were used for response times and accuracy analyses. Entry
time, or the time at which an AOI was first entered, was also analyzed using a $2 \times 4 \times 2$
ANOVA, with congruency, dyad-type, and AOI as factors.

**Behavioural data**

Reaction Time. There was a main effect of congruency, with faster responses if the dyad
contained an upfix that was congruent in meaning with its face ($F(1,34) = 36.41, p<.001, \eta_p^2= .52$). There were no other significant effects (all ps>.1).

Accuracy. There was a main effect of congruency ($F(1,34) = 54.23, p<.001, \eta_p^2= .62$),
with higher accuracy for congruent trials than incongruent trials (85.0% vs 65.9%). There were
no other significant effects (all ps>.1) (see Figure 15).

![Figure 15](image)

**Figure 15. Accuracy and Reaction Times by Congruency.** Incongruent trials were responded to less accurately and slower. Error bars represent standard error of the mean.

Eye-tracking data

For the ratio of fixations between upfix and face AOIs, there was a main effect of dyad-
type ($F(3,90) = 3.39, p=.021, \eta_p^2= .10$). Pairwise comparisons did not reveal significant
differences between any two dyad types, but the main effect of dyad type was driven by the fact that dyads that contained photographic faces showed a different (i.e., less time spent on upfixes) upfix/face ratio pattern to those with cartoony faces (.54, .59 vs., .88, 1.07), suggesting that dyads with photographic faces elicited more fixations to the face part of the dyad, whereas dyads with cartoony faces elicited equivalent numbers of fixations between the two AOIs.

**Dwell time.** The same pattern of results was observed for the dwell-time ratio between upfix and face AOIs. There was a main effect of dyad-type (F(3,90) = 3.12, p=.030, \( \eta_p^2 = .09 \)). As with the fixation ratio, pairwise comparisons showed no significant difference between any two dyad types. However, photographic faces elicited more mean dwell time on the face part of a dyad than on the upfix part of the dyad (see Figure 16).

Together, the dwell time and fixation data suggest that photographic faces require more attentional resources than cartoony faces, in that dwell time and fixations are both a rough measure of attention allocation. Participants spent much less time looking at parts of the dyads which were cartoony, suggesting that the cartoony stimuli were easy to process whereas photographs took more attentional resources.
**Figure 16.** The different dyad types and their fixation/dwell time ratios. In order: face-photo/upfix-photo, face-photo/upfix-cartoon, face-cartoon/upfix-photo, face-cartoon/upfix-cartoon. Red is dwell time, and pink is fixations. Error bars represent standard error of the mean.

**Entry time.** There was a main effect of emotional congruency (F(1,29) = 5.93, p=.021, $\eta^2_p = .17$), where the first fixation entering an AOI (regardless of which AOI) was earlier for emotionally congruent than emotionally incongruent trials. There was also a main effect of AOI on entry time (F(1,29) = 17.87, p<.001, $\eta^2_p = .38$), indicating earlier first fixations to face AOIs (196.35 ms vs. 465.73 ms). Finally, for entry time, there was a three-way interaction between congruency, dyad-type, and AOI (F(3,87) = 3.52, p=.018, $\eta^2_p = .11$). Follow-up pairwise comparisons looking at the differences between congruent and incongruent stimuli revealed that, although there were overall later first fixations for upfix stimuli, this effect was reduced in emotionally congruent trials and exaggerated in emotionally incongruent trials where there was a photographic face and cartoony upfix (347.64 ms vs. 562.13 ms, p=.002) (see Figure 17). This may suggest that cartoony upfixes were seen as more informative (that is, make it easier to
respond congruent) for emotionally congruent than emotionally incongruent trials. A slower entry time means that participants found the information less meaningful for responding, as participants will first look at what will help them to respond. Of course, for this to be true would assume that participants were covertly attending to what types of images were on screen before moving their overt attention (i.e., their gaze). That is, participants would have to be able to tell whether upfixes and faces were cartoony or photographic before moving their gaze.

![Figure 17. The different dyad types and their entry times. In order: face-photo/upfix-photo, face-photo/upfix-cartoon, face-cartoon/upfix-photo, face-cartoon/upfix-cartoon. Error bars represent standard error of the mean.](image)

**Discussion**

In this experiment, I found that face stimuli always drew more fixations and dwell time to them than did the upfixes, but that this effect was exaggerated for trials where the faces were photographic compared to when they were cartoony. I also found that for entry times to the upfix part of a stimulus, participants showed a pattern specific to stimuli where there was a photographic face and a cartoony upfix. Essentially, for this stimulus type, participants would
look at the upfix part of the dyad much sooner on congruent trials and later on incongruent trials. Additionally, as predicted, face-upfix congruency effects were evident in RT and accuracy data (with congruent trials responded to more accurately and faster). However, these congruency effects were not modulated by stimulus type – and so did not fulfill predictions that cartoony/cartoony dyads would show congruency effects not present in photographic stimuli. Finally, I also found that face stimuli consistently had more fixations and dwell time directed towards them, probably due to the saliency that faces have to the visual system (Bruce & Young, 1986; Itier et al., 2011).

Together, these results suggest two things: first, that photographic stimuli are harder to process, as they require more attentional resources than cartoony faces (likely due to their increase complexity) at least as measured by dwell time and fixations. In the future, to explore comparisons between photographic faces and cartoony faces further, cartoony faces which are more complex, to better match the photographs, could be used.

Second, for trials with a photographic face and a cartoony upfix, the upfix was likely seen as more informative in congruent trials than in incongruent trials, as it was looked at sooner in congruent trials on average. This may suggest that emotional congruency has more of an effect on upfix dyads where there is a mismatch between the media used for the upfix and for the face.

To better study the effect that emotional incongruency on mismatch photo/cartoony upfix dyads have on eye-movements, in Experiment 6 I used a new set of stimuli that made rating emotional congruency more difficult, while still manipulating whether the face stimuli were cartoony or photographic.
Experiment 6
Rationale and prediction

Experiment 5 failed to show effects that would suggest a special use of cartoony images for upfix/face dyads. For instance, the congruency effect for cartoony face-upfix stimuli were the same as for photorealistic face-upfix stimuli. The present investigation examines whether this equivalency reflects something of a ceiling effect. That is, if the congruency between face and upfix is more difficult, then perhaps the processing benefit afforded by cartoony images would be revealed. To accomplish this I sought to make it harder for participants to understand whether upfix/face dyad stimuli were congruent in their meaning. As before, I hypothesized that congruency effect for cartoony images would be greater than when the dyad involved photorealistic stimuli.

Such a result would support the proposal that, because participants perceive them as belonging together better, cartoony image dyads are more useful for creating a linguistic sentence-like structure in comics than photographs. This would further reflect a difference in perception of cartoony and photographic beyond differences in low-level features.

Methods
Participants

38 undergraduate students (Mean age = 20.4, 32 identified as women, 6 as men) with normal or corrected to normal vision participated in exchange for course credit. 5 participants were excluded based on eye tracking failure, leaving 33 participants remaining in the analysis.

Stimuli

For Experiment 6, I wanted to look more closely at more difficult situations where congruency was less obvious – that is, when the meaning of an upfix is difficult to pair with the meaning of the face below it. To increase this difficulty, I created stimuli which were in between fully congruent and incongruent in meaning, so that there would be no clear, ‘correct’ answer.
This difficulty might exaggerate reaction time, accuracy, and attentional differences between having fully cartoony stimuli vs. stimuli containing photographic elements. To this end, I created a new set of upfix dyads, with the intention of making the meaning of the upfix and the face stimulus more ambiguous in emotional congruency. As with Experiment 5, a face could be photographic or cartoony, but each now showed one of 5 expressions: happy, sad, neutral, as well as blended happy/neutral and sad/neutral, resulting in 5 total facial expression types. To match these ambiguous expressions, new upfixes were also created. They could be ‘happy’ – (a sun), sad (rain clouds), neutral (a sun behind clouds), or blended happy/neutral or sad/neutral (see Figure 18). Finally, another change from Experiment 5 is that there were no more photographic upfixes; while faces could still be photographic or cartoony, upfixes were always cartoony. This allowed me to have stimuli which varied in how many cartoony elements were present while still maximizing trials.

As there were blended stimuli instead of binary congruent/incongruent stimuli, it no longer made sense to code accuracy data in a binary manner. What this means is that rather than coding data as ‘correct’ or ‘incorrect,’ data could now be coded using ‘incongruency distance.’ For instance, a happy face with a sun above it would have an incongruency distance of 0 – that is, totally congruent. The same face with a neutral upfix (sun behind clouds) would have an incongruency distance of 2, and the same face with an upfix of storm clouds would have a distance of 4 – or, maximally incongruent (see Figure 19).
Figure 18. Example of all the different emotional content that could be in the stimuli from Experiment 6.

Figure 19. Example of incongruency distance. The numbers above the head show what the incongruency distances would be, in this case holding the upfixes constant (at neutral).
Procedure:
The procedure was identical to Experiment 5. Participants would be shown an upfix dyad, and asked to respond whether the upfix and the face in the dyad were emotionally congruent, responding “congruent” or “incongruent” with a key-press. The trial would complete upon response. As with Experiment 1, eye-tracking data were also collected.

Using every possible combination of trials (5 upfixes x 5 faces x 2 face types) resulted in 50 trials. Each of these were presented twice, resulting in 100 trials. Trials were randomly presented, but each participant would see every possible combination of upfix/face twice throughout the experiment.

Results

*Behavioural data:*

**Reaction Time.** There was a main effect of incongruency distance ($F(2.72, 95.34) = 15.59$, $p<.001$, $\eta^2_p = .31$). RTs were slower when the distance was 1 than when it was 0, and then were faster with increasingly large incongruency distances above 1. This suggests that a slightly incongruent dyad was most difficult (incongruency distance 1) to resolve, fully incongruent dyads were easiest to respond to, and the other incongruency distances falling in between these difficulty extremes (see Figure 20).
**Response type.** For behavioural responses (proportion of trials where participants answered “incongruent”) there was a main effect of incongruency distance ($F(1.74, 64.21) = 149.75$, $p<.001$, $\eta^2_p = .80$). Pairwise comparisons revealed that each degree of incongruency distance was significantly different from that preceding and following it (ps<.001, except between distance 3 and 4, $p=.01$), with (unsurprisingly) a greater likelihood of answering “incongruent” as incongruency increased. There was also an interaction between stimulus type and incongruency distance ($F(4,148) = 3.98$, $p=.004$, $\eta^2_p = .10$). Follow-up comparisons revealed that this interaction was driven by a significant difference between stimulus types at incongruency distance 2, with participants more likely to answer “incongruent” for trials that had photographic faces compared to entirely cartoony dyads ($p=.002$) (see Figure 21).
**FIGURE 21. LIKELIHOOD TO RESPOND "INCONGRUENT" BASED ON WHETHER THE DYADS HAD TWO CARTOONY ELEMENTS OR A PHOTOGRAPHIC FACE AND A CARTOONY UPFIX. ERROR BARS REPRESENT STANDARD ERROR OF THE MEAN.**

Eye-tracking data:

**Dwell time.** There was a main effect of stimulus type on \( (F(1,36) = 21.34, p<.001, \eta_p^2 = .37) \). This was driven by the fact that the ratio of upfix/face dwell time was higher for trials where both parts of the dyad were cartoons vs. when there was a mismatch (.34 vs .25). This finding suggests greater attentional deployment to the upfix part of the dyad stimuli when both parts of the dyad were cartoony vs. when the face was photographic (see Figure 22).

There was also an interaction of stimulus type and incongruency distance \( (F(2.88,103.54) = 3.54, p=.019, \eta_p^2 = .09) \). Follow-up comparisons revealed that the interaction was driven by the fact that, although there was always a relatively higher upfix/face ratio (i.e., more relative attention to the upfix part of the dyad) in trials that had cartoony faces, this effect was more exaggerated at certain congruency distances: 0 (fully congruent, .325 vs. .252, \( p = .026 \)), 2 (.331
vs. .229, p=.016), and 4 (maximally incongruent) (.427 vs. .232 p=.001). Distances 1 (.349 vs. .305) and 3 (.337 vs. .305) were less dissimilar (ps>.05).
**FIGURE 22.** DWELL TIME AND FIXATIONS BY INCONGRUENCY DISTANCE. HAVING A PHOTO FACE IN A DYAD LED TO MORE DWELL TIME AND FIXATIONS TO FACES, AND THIS WAS EXAGGERATED FOR THE MOST INCONGRUENT TRIALS. ERROR BARS REPRESENT STANDARD ERROR OF THE MEAN.

**Number of fixations.** There was a similar pattern of results to those observed for dwell time. There was a main effect of stimulus type (F(1,36) = 13.78, p=.001, $\eta_p^2=.28$). This was once again driven by a higher fixation ratio for trials that were entirely cartoony vs. trials where
the face was photographic (.369 vs. 299), with more fixations deployed to the upfix AOI when
the face AOI was cartoony.

Again there was an interaction between stimulus type and incongruency distance
(F(3.03,108.99) = 3.79, p=.012, $\eta^2_p = .09$). As with dwell time, this was driven by an
exaggeration of the upfix/face ratio towards the upfix AOI at certain incongruency distances:
distance 2 (.365 vs .286, p=.034), but especially at distance 4 (total incongruency; .429 vs. .251,
p=.001). Distances 1 (.404 vs .374) and 3 (.378 vs .350) were less dissimilar (ps > .05).

Together these fixation and dwell time results suggest that, although the face always
elicits more overt attention, it does so less in obviously incongruent situations. However, this
was observed only for dyads which were entirely cartoony, and not for dyads containing a
photographic face.

Entry time. Fewer participants were included in the analysis because some participants
did not look (at all) into the upfix AOI in some conditions, which prevented data from those
participants from being analyzed. Data from 26 Participants remained in the analysis. There was
a main effect of incongruency distance (F(4,100) = 2.63, p=.039, $\eta^2_p = .10$). Although no pairwise
comparisons reached significance between any incongruencies, the difference was greatest at
incongruency distance 4 (totally incongruent) where entry time was fastest (329.62 ms). In
contrast, entry times at incongruency distances 0-3 were within a similar range (387- 397ms).
Thus, for maximally incongruent trials, participants were faster to enter an AOI – regardless of
which AOI that was. There was also a main effect of AOI (F(1,25) = 78.91, p<.001, $\eta^2_p = .76$),
indicating sooner entry to face than upfix AOIs (128.69 ms vs. 633.59 ms), which confirmed
results from Experiment 5 that the face part of an image will elicit more attention.
Discussion

The results from Experiment 6 show that, as indexed by longer response times, the hardest conditions for participants to respond to were when the upfix and face were only slightly incongruent (distance 1), whereas the easiest was when they were maximally incongruent (distance 4). This is what would be expected, as very incongruent trials should be most obvious, and the most difficult trials should be when the upfix and dyad almost, but do not quite, match.

This experiment also replicated the findings that, while more attention is always directed to the face part of a stimulus, this pattern becomes exaggerated when the face is photographic (or attenuated when it is not). This is likely due to the fact that photographic stimuli are more complex, making them harder to process. A possible alternative explanation is that photographic faces are more attentionally salient, however it seems more likely that attention is driven to the face part of the dyads, as photographic stimuli are more complex than cartoony stimuli, and it has been established that more complex stimuli require more processing (Reppa & McDougall, 2015; Taylor, 2002). Additionally, cartoony faces have also been found to be more salient than non-face stimuli (Tomalski, Csibra, & Johnson, 2009), so a ‘face-preference’ effect, if present, should show up with cartoony faces as well.

Finally, this experiment found that, when dyads were moderately incongruent (distance 2), participants were more likely to respond “congruent” to dyads if they were cartoon/cartoon rather than photo/cartoon dyads. This evidence supports the idea that the perceived congruency of a dyad’s content can vary by the type of images used to make the dyad. That is, it is the first evidence I found that supports the idea that cartoony images may underpin the use of visual linguistic schema structures (i.e., such as upfix/face dyads), as, at least in some ambiguous situations, participants find the emotion contents of dyad components to match more if they are shown as cartoony images.
Although there was some evidence for cartoon/cartoon dyads supporting linguistic schema structures, it did not take the form that I originally hypothesized. That is, the effects required that the image be ambiguous, or difficult for participants to respond to, before any congruency effects were found. This may suggest that in situations where an upfix is obviously incongruent or congruent with a face, there is no congruency advantage to having the upfixes/face be cartoony, and photographic stimuli could support such relationships as well.

General Discussion

In the present study, I investigated how responses to upfixes and face dyads are influenced by the type of media involved (cartoony or photographic) and how easy it is to detect the emotional content of the dyad. Experiment 5 showed, in accordance with the studies from Chapter 3, that cartoony media elicit less overt attention than more complex, photographic stimuli, at least as measured by dwell time and fixations. Experiment 6 additionally showed that, for trials where there was incongruency between the upfix/face emotional information, participants perceived upfix/face dyads that were entirely cartoony to be more congruent in meaning overall compared to upfixes that contained a photographic face. It may be easier to discriminate the information from cartoony images, which encourage participants to see dyads that are entirely cartoony as congruent. An alternative possibility is that the nature of the stimulus display is more congruent when all cartoony (i.e., cartoony upfix/cartoony face suggests congruency) which biased participants to respond ‘congruent’.

The finding from Experiment 5 showing more overt attention to faces in dyads where the faces were photographic indicates that photographs garner more attentional resources than cartoony faces. This provides additional evidence supporting the conclusions made in Chapter 3. Namely, that it is easier to glean emotional information from cartoony images than naturalistic images, and that cartoony images require less attention and time to process. There was also some
evidence that, at least in trials containing both a photographic face and a cartoony upfix, that the upfix is more informative if it is congruent with the facial expression on the face — perhaps as a simpler ‘version’ of the information shown on the cartoony part of the dyad. In other words, for certain types of trials, the easy to interpret cartoony section of the dyad may represent a convenient place to rereference the emotional content of the overall image.

Experiment 6 provided more nuanced evidence. First, the eye-tracking data replicated the effect shown in Experiment 5: that dyads with a photographic face elicited more overt attention to the face. However, the behavioural results showed that, for dyads that are emotionally ambiguous, participants were more likely to respond “congruent” if both parts of the dyad were cartoony (compared to when the face was photographic), but only when the incongruency distance of 2. This suggests that while cartoony images are easier to process overall, there may be other factors informing the use of unique imagery such as upfixes in comic media.

Essentially, participants could have a schema for how an upfix informs a facial expression within a comic image, and this schema works less well if both aspects of the dyad are not cartoony. That is, just as there is evidence for an underlying structure to comic images (Cohn, 2013b), there may be an underlying understanding of how an upfix relates to a face, even outside of comics, possibly learned through culture much like comic framing and symbology (Cohn et al., 2005). If upfixes are an example of comic structure learned through cultural exposure, it may seem more appropriate when the stimuli are cartoony – as that is where participants would be exposed to the comic structure in the past. Future studies could expand on this idea.

These findings expand on previous studies showing that cartoony imagery is processed differently from photographic stimuli (e.g., Rosset et al. 2008, Churches et al. 2014). For example, while I have demonstrated that the low level features associated with cartoony images
allow for faster processing and emotion discrimination in faces (Kendall et al. 2016), the present study also suggests that this finding is not unique to faces. Rather, the effect may be more general, as other categories of cartoony images show processing advantages, such as upfixes. The eye-tracking data revealed that, as hypothesized, the presence of photographic elements in a dyad required more attentional allocation to that photographic element than to a cartoony equivalent. Along with my behavioral results, this finding suggests that cartoony imagery serves to provide similar information to that provided by photographic stimuli, only much faster. These results hold despite the fact that the current study had colour upfixes and greyscale faces. It is well-known that colour is attentionally salient (Folk, Remington, and Johnston 1992), which could have resulted in more dwell-time or fixations to the upfix stimuli than to the face stimuli. Nevertheless, the face stimuli always drew attention for longer, as well as sooner in a trial. This suggests that the saliency of the upfix stimuli did not overwhelm that of the face stimuli, because across conditions faces remained more attentionally salient, even with a shifted ratio of face to upfix fixations.

Previous work has implied that upfixes draw on a combinatorial schema because they hold a privileged relationship with their face across dimensions of location (above, not beside the head) and agreement with the face (certain expressions match certain upfixes) (Cohn, Murthy, and Foulsham 2016). Such relations hold even with novel, unconventionalized upfixes (e.g., a four-leaf clover upfix). The present data suggest that this link is strongest when both upfixes and faces use cartoony representations, since both belong to a conventionalized system of representation. That is, they may be stored in memory as systematic schematic forms within the lexicon of “visual languages”, and indeed, together comprise an abstract “upfix schema.” However, this is critically only true in situations where the relationship between the upfix and the
face is ambiguous. In most situations that I tested, photographic stimuli supported the upfix/face schema as well as cartoony stimuli did.

Photographic styles may disrupt such an expected schematic relationship. In some sense, photographic upfixes (as used in Experiment 5) or faces add another level of non-conventionalization to this overall representation. Though constraints on upfixes apply to novel, unconventional upfixes, they are still deemed as less comprehensible than familiar, conventionalized ones (Cohn, Murthy, and Foulsham 2016). Similarly, previous work has shown that upfixes are modulated by their prevalence in comics (Newton 1985). Photographic elements would provide a different type of unfamiliarity, applied to the stylistic representation of these elements rather than their conceptual content. Note that such an interpretation should not hold if upfixes used only a process of semantic construal between face and upfix (Bateman and Wildfeuer 2014)—which should not discriminate on style—rather than a schematic form stored in memory.

In sum, this work has examined the balance of cartoony and realistic representations within the context of combinatorial visual morphology of upfixes, and was thereby the first to examine eye movements in relation to these types of combinatorial visual representations. I found that participants first fixated on faces within the face-upfix dyads, but cartoony components, when present, were consistently focused on for less time, and drew fewer fixations, than photorealistic imagery. Such results indicated that, in line with prior work (Kendall et al. 2016), cartoony images are processed more efficiently than photorealistic images. These findings show that, though graphic images may predominantly be iconic, variance of aspects of systematicity and visual simplicity further modulate how they are processed. This is particularly
important within the context of visual affixation like upfixes, because the representation of face and/or upfix provides access to the entrenched schema operating to combine such parts.
Tables
Table 5 (Experiment 5)

<table>
<thead>
<tr>
<th>Dwell times</th>
<th>Fixations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Congruent</td>
<td></td>
</tr>
<tr>
<td>FP/UP</td>
<td>.548</td>
</tr>
<tr>
<td>FP/UC</td>
<td>.624</td>
</tr>
<tr>
<td>FC/UP</td>
<td>.865</td>
</tr>
<tr>
<td>FC/UC</td>
<td>1.029</td>
</tr>
<tr>
<td>Incongr.</td>
<td></td>
</tr>
<tr>
<td>FP/UP</td>
<td>.536</td>
</tr>
<tr>
<td>FP/UC</td>
<td>.552</td>
</tr>
<tr>
<td>FC/UP</td>
<td>.893</td>
</tr>
<tr>
<td>FC/UC</td>
<td>1.109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RTs</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>congr</td>
<td>Mean</td>
</tr>
<tr>
<td>Congruent</td>
<td>FP/UP</td>
</tr>
<tr>
<td></td>
<td>FP/UC</td>
</tr>
<tr>
<td></td>
<td>FC/UP</td>
</tr>
<tr>
<td></td>
<td>FC/UC</td>
</tr>
<tr>
<td>Incongr.</td>
<td>FP/UP</td>
</tr>
<tr>
<td></td>
<td>FP/UC</td>
</tr>
<tr>
<td></td>
<td>FC/UP</td>
</tr>
<tr>
<td></td>
<td>FC/UC</td>
</tr>
</tbody>
</table>
Table 6. (Experiment 6.)

<table>
<thead>
<tr>
<th>Incon. Distance</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Mean</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both cartoons</td>
<td>1827.559</td>
<td>91.303</td>
<td>.20</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>1965.210</td>
<td>110.789</td>
<td>.31</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td>2004.749</td>
<td>138.767</td>
<td>.53</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td>1877.708</td>
<td>97.256</td>
<td>.82</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>1452.359</td>
<td>73.434</td>
<td>.92</td>
<td>.031</td>
</tr>
<tr>
<td>Photo face</td>
<td>1812.861</td>
<td>91.228</td>
<td>.16</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td>2291.482</td>
<td>169.664</td>
<td>.31</td>
<td>.029</td>
</tr>
<tr>
<td></td>
<td>1864.343</td>
<td>87.257</td>
<td>.62</td>
<td>.029</td>
</tr>
<tr>
<td></td>
<td>1746.200</td>
<td>80.791</td>
<td>.85</td>
<td>.029</td>
</tr>
<tr>
<td></td>
<td>1454.954</td>
<td>79.746</td>
<td>.90</td>
<td>.032</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incongruency distance</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Mean</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congruent</td>
<td>.363</td>
<td>.053</td>
<td>.393</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>.349</td>
<td>.044</td>
<td>.404</td>
<td>.043</td>
</tr>
<tr>
<td></td>
<td>.370</td>
<td>.064</td>
<td>.407</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>.337</td>
<td>.052</td>
<td>.378</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>.477</td>
<td>.089</td>
<td>.479</td>
<td>.066</td>
</tr>
<tr>
<td>Incongruent</td>
<td>.281</td>
<td>.048</td>
<td>.340</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>.305</td>
<td>.053</td>
<td>.374</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>.255</td>
<td>.044</td>
<td>.319</td>
<td>.049</td>
</tr>
<tr>
<td></td>
<td>.305</td>
<td>.046</td>
<td>.350</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>.256</td>
<td>.059</td>
<td>.276</td>
<td>.051</td>
</tr>
</tbody>
</table>
Chapter 5: Symbols on faces

Introduction

In this chapter, I examine whether cartoony faces that contain symbols are processed differently from iconic faces, before and after the meaning of the symbols has been learned. To this end, I use the N170, an ERP, to index differences of face-sensitive brain activation before and after participants learn what a symbol standing in for the mouth means.

Outside of face perception research, there is evidence that one of the defining features of cartoony images is that they contain symbolic information. One example of this is upfixes, the symbols that have appeared on and around faces in comic media since its inception (e.g., Cohn, Murthy, & Foulsham, 2016a; Cohn & Murthy, 2013; Eerden, 2009; Forceville, 2011; Walker, 1980). Outside of comics, anime character faces have symbolic elements that change the meaning of expressions, and symbolic elements that can even replace iconic elements. There is also ongoing research cataloguing the symbols that appear in and around cartoony faces. For example, there are various descriptions of the types of symbolic images that are used to express emotion in various comics (Eerden, 2009; Forceville, 2005, 2011; Walker, 1980). There are also psychology studies on the constraints on how upfixes can be used in comic communication (Cohn & Murthy, 2013; Cohn et al., 2016a). Finally, there are studies on how different symbols are learned cross-culturally on cartoony faces to signify different emotions (Akai, Yamashita, & Matsushita, 2015; Cohn & Ehly, 2016; McCloud, 1994). However, this body of research has been largely divorced from the extensive body of cognitive psychology research on face perception.
On the other hand, although cartoony faces are commonly used as stimuli in face perception research, this body of research has not addressed the features of cartoony faces that differentiate them from photographs of faces. Despite the prevalence of cartoony faces containing symbols in comic media and animation, they are more commonly discussed within psychology in terms of how they can be efficiently iconic, and are purposefully used for that iconicity. For example, cartoony faces (or “schematic” faces) are often used in research on face perception in lieu of photorealistic faces. This is done for two reasons. First, it is easier to standardize expressions and low-level stimulus features in cartoony faces, to avoid visual confounds in experiments. Second, there is a common assumption that the defining characteristic of a cartoony face is that it is effective at being an iconic sign. That is, a cartoony face represents the characteristics of photorealistic faces through visual resemblance, and so will be perceived and processed as if it is a photorealistic face. Furthermore, research across several disciplines supports the interpretation of cartoony faces as being useful and effective because they resemble photorealistic faces. Cartoony faces are processed by overlapping brain systems and show the same EEG markers as photorealistic faces, but are more advantageous for use in several domains because they have clearer or more exaggerated features (Ferreira, Noble, & Biddle, 2006; Kendall, et al., 2016; Sagiv & Bentin, 2001).

The idea that cartoony faces are perceived in a similar manner to photorealistic faces is also supported by findings that face-sensitive areas in our perceptual systems treat iconic faces similarly to photorealistic faces (i.e., evoke activity in face-selective neural patches and have similar N170 activation) (Churches et al., 2014; Easterbrook, Kisilevsky, Muir, & Laplante, 1999; Henderson et al., 2003; Krombholz, Schaefer, & Boucsein, 2007; Latinus & Taylor, 2006; Sagiv & Bentin, 2001). However, if the purely iconic cartoony faces that are used in psychology
studies could also include symbolic elements (as with all other cartoony media), it raises questions into how cartoony/schematic faces truly function as equivalent to photorealistic faces, which almost never have symbolic elements; or whether, if they were constructed, photographic ‘symbols’ could be learned and used as easily as cartoony symbols.

This chapter seeks to answer whether a cartoony face containing a novel symbolic depiction of emotion can evoke patterns of face-sensitive brain activity in the same way as an iconic face after the meaning of the symbol is revealed to a person. That is, can a face with a novel symbolic expression be processed as if it is a known real-world facial expression? If so, is there a difference in this symbolic processing between cartoony stimuli and photorealistic stimuli? Critically, in a first study I probe the first question in a manipulation using only cartoony faces. In a second study, I probe the second question by manipulating both cartoony and photographic faces.

One way to examine the degree to which faces are processed as “face-like” is by examining a face-sensitive ERP, the N170. The N170 can be used to index the difference in face-sensitive neural responses evoked from a symbolic cartoony face and an iconic cartoony face. Although its generators and exact functions have been contested, it is robustly sensitive to the difference between a face stimulus and a non-face stimulus – e.g., a house or a flower will have a lower N170 amplitude than a face (Bentin et al., 1996; Desjardins & Segalowitz, 2013; Eimer, 2011; Itier & Taylor, 2004b; Itier et al., 2011). Studies mapping the N170 have indicated that non-faces, scrambled faces, and sketched faces (the latter two of which contain visual similarities to photorealistic faces) all show lower N170 amplitudes compared to photorealistic faces, as well as lower amplitudes compared to schematic (cartoony) faces (Bentin et al., 1996; Sagiv &
These data suggest that the N170 can discriminate a spectrum of less face-like to more face-like stimuli, and not simply face versus not-face discriminations.

In Experiment 7, I test whether our perceptual systems can learn to treat a partially symbolic cartoon-like face as equivalent to a fully iconic cartoon-like face by comparing the N170 elicited by cartoon-like iconic emotional faces with cartoon-like symbolic emotional faces in a pre-post design. Experiment 7 employs two types of face stimuli: cartoon-like faces containing iconic expressions and cartoon-like faces containing symbolic expressions. The two competing hypotheses are: 1) Even after learning the meaning of the symbolic/new expressions on cartoon-like images, there is still a difference in N170 magnitude compared to the fully iconic faces. Specifically, the symbolic expressions will have a lower amplitude N170 even after being endowed with emotional meaning. Such a result would indicate that, even though :# may mean “happy face,” that symbol is still not perceived as face-like as ☻. 2) Alternatively, after learning the meaning of the new expressions, the N170 amplitude for the symbolic face may increase to an equivalent amplitude as the fully iconic faces- indicating that depictions of expressions on faces in general may be easily learned or relearned. If our visual systems can be trained to readily perceive symbols as facial expressions, rendering the image face-like, a further question concerns whether this flexibility is unique to perception of cartoon-like faces.

In Experiment 8, I extend the results from Experiment 7 for a direct comparison between cartoon-like and photographic faces, using both novel expressions as well as canonical facial expressions. That is, I test whether the pattern of results from Experiment 7 matches the photographic stimuli, or whether people are less able to learn to perceive photographs with unknown, symbolic expressions to be as face-like as cartoon-like stimuli.
Thus, this chapter uses iconic facial expressions on cartoony faces (e.g., :(, :, :O), and compares them to cartoony faces with symbolic facial expressions (e.g., :#, :!, :H). Then, to match the cartoony stimuli, a set of photographic face stimuli were also created. First, photographs with familiar (or iconic) expressions: smiling, frowning, and neutral. Second, photographic faces using previously unfamiliar (i.e., symbolic) expressions were created.

Experiment 7

Rationale and prediction

For my final two experiments, I returned to examining a fundamental feature of cartoony faces: that they are easy to manipulate into new symbols, whether for emojis or in as stimuli cognitive psychology studies. I wanted to know whether cartoony faces were uniquely suited to endowment with symbolic meaning. Essentially, the complexity of photographic faces may limit our ability to endow them with symbolic meaning, whereas the infinite variation that can be achieved easily with cartoony faces might encourage allow them to be used and perceived more symbolically.

My first step was to demonstrate that symbolic cartoony faces could be created and, with training, come to be perceived as if they were iconic faces. I predicted that if I created cartoony images with novel symbols in place of iconic expressions, I would be able to teach participants to associate them with new meaning, despite their lack of iconicity. If a face is seen as less face-like initially because it has a random or meaningless expression, but then later it is revealed that the expression is meaningful, I would expect to find an initially reduced N170 for the novel relative to canonical facial expressions when they were meaningless (i.e., a less ‘face-like’ expression); however, after training the face-sensitive N170 should be larger in magnitude only for the novel expression faces (i.e., they would become as ‘face-like’ as the fully iconic faces).
This is because I will have altered the understanding of the novel face to be more meaningful – and so more like an iconic face. This would show that symbolic cartoony faces can take on new meaning easily, and be perceived like iconic cartoony faces, despite the lack of verisimilitude to photorealistic expressions.

An alternative pattern of results that could occur would be that the novel faces could show an attenuated N170 both before and after training; this would suggest that while the new symbols were learned, the learning did not result in the face’s becoming more ‘face-like,’ and would support the idea that symbolic faces remained differently perceived from faces that have expressions that resemble photorealistic expressions.

Subjects
25 undergraduates (19 identified themselves as female, Mean age= 20.13, 6 as male, Mean age = 18.79) provided written consent to participate in return for course credit.

Stimuli
Iconic stimuli consisted of three schematic faces canonically representing three specific emotions (happy, sad, and neutral), using simplified depictions of features of real faces.
Symbolic stimuli were face-like configurations using ASCII symbol expressions placed in the center of the face stimuli, and which held no previously established emotional meaning

All stimuli were created using Adobe Illustrator to be featurally identical except for the expression (iconic or created with the ASCII symbol) depicted in one manipulation or the presence or absence of eyes in another. All expressive features were placed in the center of the stimulus to prevent the need to move eyes from fixation. The stimuli were white and black, and displayed over a neutral grey background (RGB with 125 for all values). All stimuli were circles contained within a 500 x 500 pixel square.
The stimuli used for symbolic features were selected by a separate group of 20 undergraduate and graduate participants who rated various ASCII symbol faces on the presence of emotional information. Participants were shown 30 symbolic faces and asked to rate each for how little their expressions represented an emotion. The three symbolic expressions selected and displayed in Figure 23 are the three that most often were rated by participants as lacking associations with expressive meaning. Note that these expressions could still have meaning for participants, especially on an individual level. With the wide-spread use of symbols in media and individual differences in how people interpret emoji-like expressions, different people may understand unfamiliar expressions differently. However, these symbols were each found to be uninformative to the participants who rated them.

Control Stimuli

6 additional stimuli duplicated the iconic and symbolic images, but without eyes, to create a non-face control version of the face stimuli. These were used as control stimuli because they had the same symbols as the iconic and symbolic faces but without eyes – that is, they were the symbols without the configuration that identifies images as faces. This allowed me to compare the test stimuli and the changes they elicited in the participant’s N170 against stimuli that contained the same information without being faces.
Procedure.
At the outset of the experiment, participants were fitted with a 64 electrode Biosemi EEG cap. They were told that what they were going to be shown various faces, and that they would have to respond to what expression was shown using four keys – corresponding to happy, sad, neutral, or not an emotion. Stimuli were presented on a desktop computer using PsychoPy software (Peirce, 2007) in a sound-attenuated dimly lit room. There were three phases to the experiment: a pre-training phase, a training phase, and a post-training phase. The pre and post training phases each contained 600 trials. This was comprised of 25 blocks that displayed each of the 6 stimuli, as well as the 6 control stimuli, twice in random order. The training phase contained 15 blocks of trials. In each block, each of the six stimuli were displayed randomly three times, resulting in 18 trials per block, and 270 trials in total.
For the pre and post training phases, EEG data were recorded, in addition to RT and accuracy data. Participants were told to choose an expression label for the stimulus in each trial, but only to respond once the response screen was shown. Stimuli were preceded by a delay that was randomly varied between 900 and 1100 ms, after which the stimulus appeared in the center of the screen and was always displayed for 300 ms, during which time participants were instructed not to respond. Subsequently a response screen would be displayed, asking participants to indicate whether the emotion on the previous face was happy, sad, neutral, or not an emotion (this last option only present in the pre-phase and not the post phase). After each of the 12 images were displayed and responded to twice, participants were presented with a screen allowing them to take an additional break (see Figure 24).

In the training phase, participants were told that the previously symbolic expressions also represented happy, sad, and neutral expressions, and that they would now be expected to recognize the symbols as such. Participants were shown an instruction screen at the beginning of each run of trials. This screen assigned the previously symbolic expressions to match the meaning of the iconic expressions (e.g., :+ now became assigned as a “sad” expression). This assignment was counterbalanced across participants with every possible combination of symbolic and iconic expressions.

Unlike the pre and post training phases, the training phase emphasized performance. Participants were told to respond to the expressions displayed on the cartoony faces as quickly and as accurately as possible, and each trial ended on response. In addition, after each trial, the participants were given feedback on whether their response was correct, and how quickly they responded. Finally, in this phase I also trained participants to criterion: If there were more than
two incorrect responses in the last run of trials, the phase would continue until they reached threshold, i.e., another block of 24 trials would be repeated.

**Figure 24.** EEG was collected in the green conditions. Participants responded to each face as Sad, Happy, Neutral, or Not an Emotion. After training, all faces were assigned and so there were no more "Not an Emotion" faces.

The different design of the training phase compared to the pre and post phases was intended to allow clean ERP collection for the pre and post training phases, as well as longer presentation times to facilitate measurement of clean ERPs. The training phase was intended to ensure that participants were trained to understand the new associations between the symbolic expressions and their assigned meaning. To this end, they were encouraged to respond as quickly as possible, and were also given feedback for each trial on whether they were correct or incorrect and how quickly they responded. Finally, participants were debriefed about the purpose of the study and given their course credit.
**EEG data acquisition.** EEG procedures were identical for both experiments. Scalp EEG was recorded with 64 channel BioSemi Actiview system data. Continuous data was recorded at 512 Hz and band-pass filtered offline with half-amplitude cut-offs at 0.1 and 30 Hz. Additional electrodes were placed at the left and right mastoid processes, as well as the outer canthi of the eyes, as well as below the right eye. These additional electrodes were to record eye movements and blinks to aid in determining which trials should be discarded. All data were referenced to the average common reference.

Pre-processing was performed using the ERPlab Matlab toolbox (Lopez-Calderon & Luck, 2014). Continuous EEG data was recorded at a sampling rate of 512 Hz, and was band-pass filtered using an IIR Butterworth filter with half-amplitude cut-offs at 30 and 0.1 Hz during offline preprocessing. All data were referenced to the average of all available electrodes.

Data were segmented into epochs of 300 ms in length, which were time-locked to stimulus onset in the pre and post training phases of the experiment. Each epoch was referenced to 200 ms pre-stimulus onset. Data were processed automatically for artifacts using a moving window peak-to-peak method using a window of 100ms. This resulted in fewer than 10% of trials discarded.

The stimulus-locked EEG data from each stimulus condition were averaged separately for the pre-and post-training faces, based on expression category after training (e.g., iconic happy pre, iconic happy post, symbolic happy pre, symbolic happy post, etc.). Peak N170 activation was extracted from the epochs using a negative local peak amplitude and latency measurement with a window of 160-220ms after stimulus onset. If no reliable peak was found using this method, the participant was discarded, resulting in exclusion of one participant. Data collection was unsuccessful for an additional participant, due to peaks that were recorded well outside the
normal amplitude range, leaving 23 remaining participants. All N170 data were collected from the P9 and P10 electrodes as the mean N170 signal across all conditions was strongest and most reliable at these sites. The N170 is often reported to be maximal at six occipital-temporal sites on a 64 channel system (P7, P9, P07, P8, P10, P08); however, it is also common to find the clearest N170 effect at P9/P10, which are the most ventral of the six sites (Bentin et al., 1996).

Results

To summarize, in this experiment participants first underwent a pre-training phase in which they labeled symbolic and iconic expressions on cartoony faces. They were then retrained to see the symbolic expressions as representing the same expressions as the iconic expressions in a training phase, which involved speeded response and training to criterion. Participants then completed a post-training phase that was identical to the pre-training phase except that there was no longer a “not an emotion” response available. EEG data was collected from occipito-parietal electrode sites before and after training. The central question concerned whether the assignment of meaning to arbitrary facial expressions would evoke a corresponding shift in N170 amplitude from pre- to post-training, reflecting the rapid acquisition of neural activation patterns associated with face categorization.

Reaction time and accuracy of expression identification were analyzed for pre- and post-training phases using repeated measures ANOVAs with the following within-subject factors: phase (pre, post) x stimulus type (iconic, symbolic). For the purpose of analysis, each of the symbolic stimuli were coded based on the expression that they were assigned in the training phase. Additional and separate ANOVAs were also performed on values from control stimuli (i.e., the faces without eyes). For the training phase 2 x 3 repeated measures ANOVAs were employed with stimulus type (iconic, symbolic) and expression (happy, sad, neutral) as within-
subject factors. Finally, outliers for RTs were removed when they exceeded an absolute threshold of 5 seconds, so that fluctuations in the response times of training phases, which typically began with much slower response times, would not be biased for removal using a statistical rule.

Values were adjusted using the Huynh-Feldt adjustment for violations of the assumption of sphericity where appropriate. All pairwise comparisons are Bonferonni corrected. The full data can be found at the end of the chapter.

**Behavioral Results**

**Pre-and Post- training.**

*Reaction time.* For the pre- and post-training phases, participants were instructed not to respond right away, but rather to wait until a response screen was displayed. The overall short reaction times (M= 525 ms in pre-, M= 540 ms in post-) reflect the fact that participants knew when the response screen would occur and could prepare their responses. Reaction time results reported below were calculated using only correct trials. Results showed a main effect of phase \([F(1,23)=64.29, p<.001, \eta^2_p = .74]\), with faster responses in post than in pre, reflecting a training effect. I also observed a main effect of stimulus type \([F(1,23)=6.96, p=.015, \eta^2_p = .23]\) with slower responses to iconic relative to symbolic stimuli. This likely reflects the fact that in the pre-training block the symbolic stimuli did not have to be decoded but were simply identified as symbolic with a single key. No other effects were observed (ps>.1). Accuracy for the pre and post-training phases was generally high and is reported at the end of the chapter.

**Training Phase.**

*Accuracy.* For the training phase, there was a main effect of expression \([F(2, 48)=4.50, p=.016, \eta^2_p = .16]\), with overall lower accuracy for neutral expressions.
**Reaction Time.** For the training phase, I observed only a marginal effect of stimulus type, with faster responses to iconic than symbolic stimuli. The overall similarity between the stimulus types in the training phase suggests that participants were able to learn the new stimulus associations successfully. However, to better understand the training phase, as well as to confirm that participants were successfully learning the new meanings for my symbolic stimuli in the training phase, I compared the accuracy and reaction times for responses to the symbolic and iconic faces across blocks to see when the symbolic stimuli responses reached the threshold of the iconic responses. Results revealed slower responses for symbolic stimuli in early but not later blocks. Accuracy only differed within the first block (see Figure 25).
Figure 25. The novel expressions (in blue) initially elicited slower responses than familiar expressions, but reached the same latency range as the iconic expressions within three blocks. Accuracy only differed in block 1. Error bars represent standard error of the mean.
ERP Results

ERP amplitudes and latencies were analyzed using repeated measures ANOVAs with the following factors: phase (pre, post) x stimulus type (iconic, symbolic) x electrode (P9, P10). Because the visual dissimilarities between control and test stimuli may have obscured clear N170 results (i.e., the N170 is known to be affected by low-level features (Rossion & Caharel, 2011; Yue et al., 2011), which differ from control to test stimuli, whereas the features pre-post did not differ), test stimuli were analyzed separately.

Two additional ANOVAs were also performed on control stimuli (those without eyes). Values were adjusted using the Huynh-Feldt adjustment for violations of the assumption of sphericity where appropriate. All follow-up comparisons were Bonferonni corrected. All results are reported in Table 7.

N170 Amplitude. For N170 magnitude there were no main effects of either phase or stimulus type (ps > .2). Crucially, there was an interaction between phase and stimulus type $[F(1,22)=4.78, p=.04, \eta_p^2 = .18]$ such that N170 magnitude differed between pre and post-training for symbolic but not for iconic faces. Pairwise contrasts revealed a trend-level increase in amplitude for symbolic faces (p=.06) but not for iconic faces (p=.98). Finally, there was a three-way interaction between phase, stimulus type and site $[F(1,22)=4.55, p=.044, \eta_p^2 = .17]$. Pairwise contrasts indicated an increase in N170 amplitude after training for symbolic faces at site p9 (p=.02) but not site p10 (p > .2). In summary, N170 magnitude for symbolic face stimuli increased with training, such that after training it was more equivalent to N170 magnitude for iconic faces, and this effect was left lateralized over temperoparietal cortex (see Figure 26). As hypothesized, the N170 for symbolic expressions was magnified by assigning meaning to the
expressions during training. Finally, there was a three-way interaction between phase, stimulus type, and sensor site suggesting lateralization of the training effect for symbolic stimuli.
**Figure 26.** The magnitude for symbolic (novel) faces increased due to training (in blue). The N170 magnitude for iconic faces (in red) remained constant. Error bars represent standard error of the mean.

*N170 Latency.* For N170 latency there was a main effect of stimulus type, \([F(1,23) = 5.75, p = .05, \eta^2_p = .16]\), revealing longer latencies in response to symbolic vs. iconic stimuli. There was also a trend-level effect of phase, \([F(1,23) = 4.22, p = .05, \eta^2_p = .15]\), with overall longer latencies post-training. There was no interaction between stimulus type and phase \((p > .60)\).

**ERP Results: Control stimuli**

*N170 Magnitude.* The same ANOVAs were performed on the control stimuli as on the test stimuli analyzed above. Note that these stimuli were identical to the test stimuli save that they did not have ‘eyes’ (they were simply the expression stimuli inside of a white circle). There was a main effect of stimulus type \(F(1,22) = 51.23, p < .001, \eta^2_p = .70\), with larger overall N170
magnitudes for iconic faces (-6.69 vs. -5.25), and a main effect of sensor site [F(1,22)=4.63, 
p=.043, η²_p = .17] with larger amplitudes at p10 (-6.52 vs -5.42)(Bentin et al., 1996). No other effects were observed (ps>.05). Thus, for symbolic faces, the results of training were only observed when eyes were present, maintaining a face-like featural configuration (see Table 7).

**N170 Latency.** For the experimental stimuli, a main effect of stimulus type was found [F(1,22)=17.12, p<.001, η²_p = .44] revealing longer latencies in response to symbolic vs. iconic stimuli.

**Discussion**
These findings show that symbolic expressions elicited a similar N170 to iconic expressions, but only once they were assigned communicative meaning during the training phase. That is, training caused the pattern of N170 magnitude in response stimuli with symbolic expressions to increase to a magnitude similar to that observed for familiar iconic expressions. In contrast, the iconic expressions, predictably, did not elicit any shift in N170 magnitude after training. This effect was not observed in control stimuli, suggesting that it was dependent on a complete face configuration. Overall, these results support the hypothesis that symbolic expressions in cartoony faces are processed in the same manner as iconic expressions when a face-like configuration is present.

The results from this experiment also suggest that novel expressions on cartoony faces can be learned easily, and that after learning they will evoke an N170 response that is equivalent in amplitude to those evoked by canonical iconic faces. The ease at which the training occurred is notable – fewer than 3 blocks of 24 trials were necessary for participants to respond to the novel faces as if they were well-known – that is, both as quickly and as accurately.

In the next experiment, I introduce photographic faces to directly compare effects of learning ‘symbolic’ photographic faces in comparison to cartoony faces.
Experiment 8

Rationale and prediction

After finding that cartoony faces with novel expressions could easily be learned and become meaningful, and that after learning they showed a deflection of N170 magnitude in the direction of iconic faces, in Experiment 7, a natural follow-up question was whether this capacity to take on new information was unique to cartoony faces or whether responses to photographic faces, though harder to manipulate, could be similarly trained. Thus, the next experiment compared training effects on ERP responses to cartoony faces with novel expression information with photographic faces with novel expression information.

I predicted that novel expressions on photographs would show a different pattern of results than for cartoony faces, with the novel expressions only showing increased N170 magnitude after training for cartoony faces and not for photographic faces. That is, I predicted that photographic faces with new meaningful shapes would not be able to be learned as well as cartoony analogues. Put another way, I expected a replication of results from Experiment 7 for cartoony stimuli, but not for photographic stimuli.

Subjects

28 undergraduates (21 identified themselves as female, Mean age = 20.65, 6 as male, Mean age = 21, and 1 declined to answer, Age = 20) provided written consent to participate in return for course credit. Two participants were excluded due to having ERP amplitudes outside of the normal range, leaving a sample size of 26 participants.

Stimuli

The procedure for Experiment 8 was identical to that of Experiment 7, except that I eliminated the eyeless control stimuli. In their place, I created a new set of photographic stimuli.
These were created to be analogous to the entire set of cartoony stimuli so that a direct comparison between cartoony and photographic stimuli was possible. To this end, I employed photographic faces with six expressions: a smiling face, a frowning face, a neutral face, and then three ‘novel’ expressions that were meant to be tied to no existing emotion. For the new, photographic faces, I had one actor perform various facial expressions – first, she made happy, sad, and neutral faces to match the iconic cartoony stimuli. To create a battery of novel(symbolic faces, the actor performed as many different unique expressions as was possible for her, resulting in 30 possible facial expressions.

Thus, as with the faces chosen in Experiment 7, the photographic faces were chosen out of a battery of 30 possible symbolic faces. To select which ones would be used, each of the 30 faces was rated by 20 undergraduate and graduate university students in a pilot experiment as the ‘least emotionally meaningful’ expressions. Participants were shown all of the 30 faces and asked which of the faces had the least meaningful expressions. The three faces which were most agreed upon by participants were used (see Figure 27).

Procedure

As with Experiment 7, in the pre and post-training phases, participants were shown each type of face and asked to respond “happy,” “sad,” “neutral,” or “not an emotion” to each stimulus while EEG was being recorded (although ‘not an emotion’ was only an option in the ‘pre’ condition). As previously, in the training phase, they were told that each of the novel, “not an expression” faces were actually meaningfully the same as the happy/sad/neutral faces. This association between expression and emotion was randomly assigned and counterbalanced across participants, and as in the previous experiment participants were trained to criterion to respond to
the association correctly before being allowed to proceed to the final test phase. They then went back to rating the faces as happy/sad/neutral.
Results

Repeated measures ANOVAs included the same factors as Experiment 7: phase (pre or post), stimulus type (iconic or symbolic), and sensor location (p9 or p10) for ERP results. However, there was now an additional factor: image type (photorealistic or cartoony). All results are reported in Table 8.

Behavioral Results

Reaction time (Pre- and Post- training).

As with Experiment 7, outliers were removed from RT data using an absolute threshold of 5 seconds, to keep the integrity of the (non-outlier) slower responses in the initial training phases.
Replicating Experiment 7, there was a main effect of phase type [F(1,25) = 32.37, p<.001, \( \eta^2_p = .55 \)]. This was due to longer reaction times in the pre-training phase than in the post-training phase (772 ms vs 482 ms), reflecting a training effect. There was a main effect of stimulus type [F(1,25) = 12.60, p=.001, \( \eta^2_p = .33 \)], with slower responses to the iconic images compared to the symbolic images (651 ms vs. 602 ms) – an effect not found in Experiment 7, but which is reasonable as participants could quickly answer ‘not an expression’ to many of the symbolic faces in the pre-phase. There was also a main effect of image type [F(1,25) = 72.49, p<.001, \( \eta^2_p = .74 \)]. Cartoony faces elicited faster responses than photographic faces (520 ms vs. 734 ms).

Unlike in Experiment 7, main effects were qualified by an interaction between phase and stimulus type [F(1,25) = 79.72, p<.001, \( \eta^2_p = .75 \)]. All follow-up comparisons were significant (ps<.005), but the interaction was primarily driven by a change in the pattern between stimulus types between pre and post training: Whereas in the pre-training phase, symbolic faces elicited faster responses than iconic faces (682 ms vs. 862 ms), in the post-training phase, the iconic faces elicited faster responses than the symbolic faces (441 ms vs. 523 ms). This is logical, as all the symbolic faces could be lumped together as “not an emotion” in the pre-phase, but had to be individually labeled for specific emotion post-phase. There was also an interaction between stimulus type and image type [F(1,25) = 40.13, p<.001, \( \eta^2_p = .61 \)]. Follow up comparisons revealed that, for photographs, there was no difference in response time between iconic faces and symbolic faces (p= .371). In contrast, for cartoony faces, symbolic faces elicited faster responses than iconic faces (579 ms vs. 461 ms) (p=.001), again possibly representing that symbolic faces were easy to respond to in the pre-phase, when the same response was used for all of them. Finally, there was also an interaction between phase, stimulus type, and image type [F(1,25) =
10.90, p = .003, $\eta^2_p = .30$. All follow-up comparisons were significant (ps<.001) save for the comparison of symbolic to iconic faces for cartoony images in the post phase. Specifically, response times to cartoony faces in the post-training phase revealed a floor effect, as they were equally fast, and faster than responses to any of the other images (375 ms and 342ms for the novel and iconic cartoony images in the post phase, compared to 465 ms and 591 ms for the photographs in the post phase). All other comparisons were non-significant (ps>.05).

Together these results contribute to the body of findings indicating that cartoony images are faster and easier to interpret, even when they contain symbolic content. Cartoony images always elicited faster responses, across all conditions, and these responses were fastest following training.

**Reaction times (Training Phase)**

There was a main effect of stimulus type [$F(1,25) = 19.92, p<.001, \eta^2_p = .43$], with iconic faces again eliciting overall faster response time than symbolic faces. There was a main effect of image type [$F(1,25) = 74.07, p<.001, \eta^2_p = .73$], with cartoony faces eliciting faster average response times than photographic faces, an effect not found in Experiment 7. Finally, there was a main effect of expression type [$F(2,50) = 4.27, p=.019, \eta^2_p = .14$]. Follow-up comparisons revealed that this reflected a slightly slower average response time for neutral faces compared to happy faces (p=.021).

There was also an interaction between stimulus type and image type [$F(1,25) = 40.36, p<.001, \eta^2_p = .60$]. Follow-up comparisons revealed that this interaction was driven by a greater difference between iconic and symbolic expressions for photographs (1674 ms vs. 1370 ms)(p<.001), than between iconic and symbolic expressions for cartoony faces (1430 ms vs. 1304 ms) (p=.037). These new effects compared to Experiment 7 are likely due to the inclusion
of ‘symbolic’ photorealistic faces in this experiment, which were not present in the previous experiment (see Figure 28).

Accuracy

There was a main effect of stimulus type \([F(1,25) = 12.98, \ p=.001 , \ \eta^2_p = .33]\), with slightly lower accuracy for symbolic faces than iconic faces (89.7% vs 91.8%). There was a main effect of image type \([F(1,25) = 14.88, \ p=.001 , \ \eta^2_p = .36]\) with lower accuracy for photographs than cartoony faces (89.5% vs 92%). Both of these results were new to Experiment 8 compared to Experiment 7, and likely due to the inclusion of a new type of face stimuli, the symbolic photorealistic face. There was also a main effect of expression type \([F(2,50) = 4.22, \ p=.020, \ \eta^2_p = .14]\), as with Experiment 7, with poorer accuracy for neutral faces compared to happy faces (89.7% vs 92.4%) (p=.022). Finally, there was an interaction between stimulus type and image type \([F(1,25) = 26.47, \ p<.001, \ \eta^2_p = .50]\). Follow-up comparisons revealed that this reflected a greater accuracy difference between iconic and symbolic expressions for photographs (92.8% vs 86.7%) (p<.001), than between iconic and symbolic expressions for cartoony faces (91.3% vs 92.3%).

A few patterns emerged. First, as with Experiment 7, participants were able to learn the new associations very quickly, with accuracy and response times for symbolic faces converging with the iconic faces after only a few blocks of trials. However, this pattern was more pronounced for cartoony faces than for photographic faces – yet again confirming that emotional information conveyed by cartoony faces is easier to acquire and process than that conveyed by photographs. In fact, during training while responses to photographic symbolic faces reached equivalent accuracy to other stimulus categories, they continued to elicit the slowest response times for the entirety of the training phase.
**Figure 28.** Accuracies and response times for symbolic faces took a few blocks to converge with iconic faces. While both types of icons were responded to quickly and accurately to begin with, it took roughly 3 blocks for cartoony symbols to catch up. Photographic symbol faces were never responded to as quickly. Error bars represent standard error of the mean.
ERP Results

N170 Magnitude

There was a main effect of phase type [F(1,25) = 7.88, p=.01, \( \eta^2_p = .24 \)]. This effect was due to overall reduced N170 amplitudes from pre- to post-training (-6.23 to -5.65), likely due to habituation to the face stimuli. There was a main effect of image type [F(1,25) = 4.47, p=.045, \( \eta^2_p = .15 \]): Photographs elicited smaller N170s than cartoony faces (-5.16 vs -6.73). This replicates the differences between photographic and cartoony faces found in the ERP study from Chapter 3 (see Figure 29).

These main effects were qualified by an interaction between phase and stimulus type [F(1,25) = 4.73, p=.01, \( \eta^2_p = .16 \)]. While iconic and symbolic faces both elicited reduced N170s from pre- to post-training, this reduction for pre- to post- training was more dramatic for iconic faces than symbolic faces (as measured by comparing the iconic vs. symbolic faces in the post phase p= .002). As in Experiment 7, regardless of whether the image was photographic or cartoony, the change in N170 amplitude from pre- to post-training differed depended on whether the expression was canonical (iconic) or newly learned (symbolic). The finding that the overall reduction in amplitude over time was reduced for symbolic faces suggests that there is less habituation for symbolic faces because, through learning, the cartoony face has become something new, reducing habituation. Although the stimuli remain the same visually, their meaning has changed. These data suggest that the flexible capacity of the visual system to ‘learn’ novel expressions is not specific to cartoony images. However, it is also worth noting, that of all the types of images shown to participants, only cartoony symbolic stimuli showed an increase in N170 magnitude from the pre- to post-training (-6.62 to -6.85), while all other types of images showed an N170 reduction. This pattern more closely mirrors the results from Experiment 7, but
was not statistically significant — potentially because this study was not sufficiently powered to detect 3-way interactions.

There was an interaction between phase and image type \([F(1,25) = 25.16, p<.001, \eta_p^2 = .50]\). Follow-up comparisons revealed that this was due to a reduction in N170 pre- to post-training for photographic face \((p<.001)\), but that this reduction was not observed for cartoony faces \((p=.903)\) (see Figure 30 and Figure 31).
**Figure 29.** Cartoony faces showed less of a N170 decrease pre-post compared to photographic faces. These results are not an effect of P1 differences, as described in the results.
**Figure 30.** On the top, difference scores P1-N170, revealing the same effects as the raw N170 scores. On the left, cartoon vs. photographic faces. On the right, iconic vs. symbolic faces. Cartoon faces did not elicit an attenuated N170 pre-post, whereas photographic faces did. On the right, iconic faces showed more of an N170 attenuation pre-post compared to symbolic faces. Error bars represent standard error of the mean.
**Figure 31. Symbolic faces showed less of an N170 decrease compared to iconic faces. Again, these results cannot be explained as continuation of P1 patterns.**

*N170 Latency*

There was a main effect of phase type \([F(1,25) = 7.79, p=.01, \eta^{2}_p = .24]\). On average, N170s latencies were longer in the post-training phase than the pre-training phase (163ms vs 159ms). There was a main effect of expression \([F(2,50) = 3.24, p=.047, \eta^{2}_p = .12]\). Although follow-up comparisons were not significant, this is likely due to neutral faces eliciting a later N170 compared to the other expressions (163ms vs 159ms – for both). There was a main effect of image type \([F(1,25) = 34.46, p<.001, \eta^{2}_p = .58]\). Photographs elicited an N170 sooner than cartoony images (147ms vs. 175ms). There was a main effect of sensor site \([F(1,25) = 9.65, p=.005, \eta^{2}_p = .28]\). N170s from p9 were slightly later than at the p10 sensor (164ms vs. 157ms).
There was an interaction of expression and image type \[F(2, 50) = 3.62, \ p=.034, \ \eta^2_p = .13\]. Although the follow-up comparison did not reach significance (\(p=.069\)), it was due to neutral cartoony faces showing a later N170 than other image types (180ms). There was also an interaction of expression and sensor site \[F(2, 50) = 4.72, \ p=.013, \ \eta^2_p = .16\]. Follow-up comparisons revealed that this was due to cartoony neutral images eliciting a later N170 than happy faces (\(p=.012\)) and sad faces (\(p=.037\)), but only at the p9 sensor. Finally, there was an interaction of image type and sensor site \[F(1, 25) = 7.57, \ p=.011, \ \eta^2_p = .23\]. There was no difference between sensors for photographic images, but cartoony images showed a later N170 at the p9 sensor than the p10 sensor (\(p=.003\)). This may be evidence in favor of the idea that the networks generating the N170 are more sensitive to photorealistic faces, if not photorealistic expressions.

**Peak to peak comparison**

With N170 results, it is important to control for the possibility that amplitude differences reflect differences in the preceding P1. That is, a difference in P1 amplitudes may persist into the N170, and spuriously appear to be N170 effects. To control for effects of P1 amplitude, I used a peak-peak analysis comparing the difference between the P1 peak and the N170 peak. If the same pattern of results is observed the analysis serves as confirmation that the effects reported above are indeed N170 effects.

Peak to peak analysis revealed a main effect of image type \[F(1, 25) = 7.82, \ p=.010, \ \eta^2_p = .24\]: This effect is consistent with the N170 data reported above; the P1-N170 difference for photographs was smaller than for cartoony images (8.58 vs 10.64).

There was also an interaction between phase and stimulus type \[F(1, 25) = 5.53, \ p=.027, \ \eta^2_p = .18\]. Follow-up comparisons were not significant, but as with the N170 results, the P1-
N170 difference for iconic faces was attenuated pre-post, whereas the symbolic face P1-N170 increased pre-post.

There was an interaction between phase and image type \([F(1,25) = 32.62, p<.001, \eta^2_p = .57]\). Once again, this reflects the pattern reported in the N170 results. The P1-N170 difference remained consistent pre and post for cartoony images, but was reduced for photographic images \((p=.001)\).

Together, results of this control analysis confirm that the earlier N170 results were, reported above did not merely reflect P1 differences. The exception to this is that when controlling for the P1, there was no main effect of phase type, suggesting the overall habituation found in the N170 results was actually a habituation effect on the P1 from pre-post training, although it is important to note that the more specific habituation difference between photographic faces and cartoony faces from pre- to post- training phase was found on the N170, and cannot be explained by P1 differences.

Discussion
This experiment built upon the findings from Experiment 7. First, I found that, as with cartoony faces, novel expressions on photographic faces could be learned quickly and easily, although there was a behavioural advantage in learning new expressions on cartoony faces. I also found no magnitude difference between the two types of stimuli from pre- to post-training, suggesting that the N170 for photographs and cartoony images was equally influenced by learning.

One of the more notable results from this study is that, whereas I observed an N170 habituation effect for ERP responses to photographic faces, which is to be expected from face stimuli which are presented many times (Heisz, Watter, & Shedden, 2006b; Leinbach & Fagot, 1993), habituation was not observed for cartoony faces. This, along with the increased magnitude
for cartoony faces from pre- to post- training may suggest that they did indeed become more ‘face-like’ from having new meaning assigned to them, or that they were perceived as more novel when expressions were learned compared to the photorealistic faces.
General Discussion

The present chapter examined whether learning the meaning of initially meaningless, partially symbolic, cartoony expressions influenced the magnitude of N170 elicited by those stimuli compared to canonical facial expressions. It also tested whether any such effects were unique to cartoony imagery or could be found when we created novel expressions on photographic faces as well. To do so I used a pre-post design where participants were initially presented with canonical facial expressions (iconic faces), as well as faces that had mouth configurations that did not resemble configurations of familiar facial expressions (symbolic faces). Half-way through each experiment, participants learned that the novel configurations represented meaningful emotional expressions, which matched the emotions expressed by the iconic face stimuli. N170 response to both types of stimuli were compared before and after training of emotional associations with the novel symbolic stimuli.

In Experiment 7 I found that, while the symbolic stimuli elicited lower amplitude N170s than fully iconic stimuli pre-training, N170 amplitudes increased for symbolic stimuli such that they were equivalent to those elicited by iconic stimuli after training. However, for this effect to be found, cartoony faces still required eyes – that is, the configuration of the face stimulus was necessary to find the N170 effects that resulted from the training phase. In Experiment 8, I found evidence that the difference between symbolic and iconic stimuli was not specific to cartoony images, but applied to photographic images as well. However, there was still a notable difference between photographic and cartoony stimuli. Here, whereas the N170 elicited by photographic faces was attenuated, cartoony stimuli elicited no N170 magnitude reduction pre- to post-training. The results from Experiment 8, taken together, suggest that novel expressions on faces can be learned regardless of the degree of photorealism of the face. In addition, while N170s to
emotionally expressive photographic faces evoked habituation effects with repeated presentation, emotionally expressive cartoony faces did not.

Although it may seem intuitive that there would be a change in face-sensitive neurophysiological responses to stimuli when they are imbued with meaning, these two experiments also provided an opportunity to examine any differences in the flexibility with which we can endow cartoony images with new meaning compared to photorealistic images. Cartoony images commonly contain symbols in place of iconic elements, and so it would be reasonable to assume that the cartoony images would facilitate learning based on a mix of symbols and icons. This is not what I found – unfamiliar, or unassigned expressions (i.e., expressions which were symbolic) were easily learned, as measured by RTs and accuracy, for both cartoony images and photographs, suggesting that there is no unique cartoony advantage for symbolic information.

Even though there does not seem to be an advantage for cartoony symbolic information, it seems likely that the processing advantage previously found for cartoony images as a whole also applies here. That is, cartoony faces tend to contain overtly symbolic stimuli because their ease of processing allows more information to be added to them. This quality of cartoony imagery may shed light on why cartoony faces, compared to photographic faces, are more easily interpretable to people with autism (Rosset et al., 2008) and why emojis are used so widely to communicate emotion information (Hakami, 2017; Moschini, 2016; Yuasa, Saito, & Mukawa, 2006). In short, they simply can be processed fast enough that extra information is not a perceptual burden.

One striking aspect of this study is in the behavioral results. As would be expected, when the unassigned symbols were first assigned meaning, there was a difference in response time
between the iconic and symbolic stimuli during the learning phase, with the iconic stimuli being responded to more rapidly. What is notable however, is how quickly this difference between the types of stimuli was resolved. On average, and in both experiments, by the end of three blocks of trials, the symbolic stimuli were being responded to just as fast and as accurately as the iconic stimuli, which represents only 72 trials, and 1/5th of the entire learning block. The rapidness at which participants were able to learn the new assignments of the symbolic stimuli suggests that symbols are not only able to be perceived as part of faces, but that this inclusion is relatively easy to do. This finding comes with a caveat, however. Experiment 8 yielded behavioural results that showed it was more difficult to learn new expressions on photographic faces compared to cartoony faces. Although participants were able to become equally accurate for photos and cartoony images, they were also always a little slower in responding to photographic images. This finding builds on the previous studies in this dissertation, suggesting that the unique quality of cartoony images that encourages symbols to be added may simply be that they are easier images to learn and process overall.

Experiment 8 showed interesting patterns based on habituation effects. Finding differences in habituation is interesting because, although habituation of the N170 based on repeated exposure to faces is common (Heisz et al., 2006b; Maurer, 2008), by definition habituation effects are not observed for novel stimuli (Caharel, d’Arripe, Ramon, Jacques, & Rossion, 2009; Heisz et al., 2006b). N170 habituation effects are also not observed for inverted faces (Heisz, Watter, & Shedden, 2006a), and amongst people who score highly in autism traits (Webb et al., 2010). That is to say – habituation occurs in the N170 for familiar faces but not in novel faces, or in non-faces. This would suggest that, from my data, the novelty of the symbols gaining new meaning ‘resets’ the habituation that the symbolic faces would otherwise show. The
cartoony faces may be akin to inverted faces – containing expression information without truly being perceived as faces, or more accurately, activating parts of the underlying components of the N170 face-sensitive ERP while not activating others. Conceptually interesting in this comparison, however, is that while inverted faces are typically difficult to process (Chien, 2011), cartoony images are still processed readily. However, schematic faces have been used in the past successfully for N170 habituation studies (Easterbrook et al., 1999; Leinbach & Fagot, 1993), which might represent that a certain level of detail is necessary for N170 habituation to occur, and so N170 face habituation may be a result of underlying neural patches which respond only to detailed, photographic faces, while other components that comprise the N170 continue to respond to both types of faces. At least one interesting analogue to the stimuli from my studies appears in the literature: a study found that non-human robotic faces still showed N170 differences based on expression but not the same inversion effects as human faces elicited. The authors concluded that humans can read expressions on faces that are not necessarily realistic human faces (Chammat, Foucher, Nadel, & Dubal, 2010). Cartoony faces may be like those robotic faces, expressing emotions without being perceived as human faces.

Past research, such as that reported in Chapter 3, has cast some doubt on whether even an iconic cartoony image is usefully comparable to a photorealistic image when trying to communicate information, suggesting that iconic cartoony images elicit larger N170 amplitudes and faster latencies than photo-realistic stimuli. While some cognitive psychology studies use cartoony images in place of photorealistic images in order to better control stimulus differences, based on the results of these and the studies from earlier chapters (e.g., Chapters 3 and 4), it increasingly seems that cartoony faces are processed quite differently from photorealistic faces, at least when it comes to discriminating between expressions, and when the cartoony faces are
more simplified. My data suggest that this difference in processing is not due to some qualitative difference, but simply due to the level of an image’s simplicity or low level features, or how easily these features allow symbolism to be learned. Greater care should then be taken to ensure that cartoony images are iconic if they are used in a context meant to be directly analogous to photorealistic stimuli, and even then with the caveat that it is possible that no photograph will ever show as strong effects (Ferreira et al., 2006; Joy Lo, Yien, & Chen, 2016; Setlur, Albrecht-Buehler, Gooch, Rossoff, & Gooch, 2005)

This study may also help with understanding when to include symbols in non-face images, such as computer icons or in pictorial instructions. Past research has taken some strides in understanding the use of cartoony images rather than text to disseminate important information cross-culturally where language (and translation) is a barrier, such as for learning and recognizing information within medical, military or traffic contexts (Clawson, Leafman, Nehrenz, & Kimmer, 2012; Fierro, Gómez-Talegón, & Alvarez, 2013; Tack et al., 2014). If learning information contained cartoony images is easier than learning that contained in photographs, and the information in them remains more easily transmitted after learning, it supports their use in contexts where rapid recognition is necessary. Additionally, if symbols can easily be integrated with icons in cartoony imagery, there may be a way to create more sophisticated iconography for the use in cross-cultural initiatives, such as representing the symptoms of disease without the need for true translation (Green & Myers, 2010).
### Table 7 (Experiment 7)

<table>
<thead>
<tr>
<th></th>
<th>test stimuli</th>
<th>control stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre</td>
<td>post</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Icons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td>p9</td>
<td>-5.384</td>
</tr>
<tr>
<td></td>
<td>p10</td>
<td>-5.869</td>
</tr>
<tr>
<td>happy</td>
<td>p9</td>
<td>-5.272</td>
</tr>
<tr>
<td>sad</td>
<td>p9</td>
<td>-6.373</td>
</tr>
<tr>
<td></td>
<td>p10</td>
<td>-7.608</td>
</tr>
<tr>
<td>Symbols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td>p9</td>
<td>-5.226</td>
</tr>
<tr>
<td></td>
<td>p10</td>
<td>-6.233</td>
</tr>
<tr>
<td>happy</td>
<td>p9</td>
<td>-5.042</td>
</tr>
<tr>
<td>sad</td>
<td>p9</td>
<td>-5.275</td>
</tr>
<tr>
<td></td>
<td>p10</td>
<td>-6.194</td>
</tr>
</tbody>
</table>
### Table 8 (Experiment 8)

<table>
<thead>
<tr>
<th></th>
<th>Icons</th>
<th>neutral</th>
<th>p9</th>
<th>Mean</th>
<th>Std. Error</th>
<th>post</th>
<th>Mean</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cartoons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>happy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>happy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Symbols</td>
<td>neutral</td>
<td>p9</td>
<td>-6.517</td>
<td>.958</td>
<td>-6.901</td>
<td>1.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>happy</td>
<td>p9</td>
<td>-6.418</td>
<td>.942</td>
<td>-6.674</td>
<td>.871</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sad</td>
<td>p9</td>
<td>-7.223</td>
<td>1.036</td>
<td>-7.009</td>
<td>.991</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p10</td>
<td>-7.678</td>
<td>.874</td>
<td>-6.641</td>
<td>.868</td>
</tr>
<tr>
<td></td>
<td></td>
<td>happy</td>
<td></td>
<td>p9</td>
<td>-6.271</td>
<td>.999</td>
<td>-7.135</td>
<td>1.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sad</td>
<td>p9</td>
<td>-6.742</td>
<td>.911</td>
<td>-6.676</td>
<td>.889</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p10</td>
<td>-7.228</td>
<td>.769</td>
<td>-7.137</td>
<td>.882</td>
</tr>
<tr>
<td></td>
<td></td>
<td>happy</td>
<td></td>
<td>p9</td>
<td>-6.581</td>
<td>.720</td>
<td>-6.555</td>
<td>.749</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sad</td>
<td>p9</td>
<td>-6.794</td>
<td>.726</td>
<td>-7.109</td>
<td>.769</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p10</td>
<td>-7.728</td>
<td>.769</td>
<td>-7.137</td>
<td>.882</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Symbols</td>
<td>neutral</td>
<td>p9</td>
<td>-5.836</td>
<td>.917</td>
<td>-4.738</td>
<td>.848</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>happy</td>
<td>p9</td>
<td>-6.387</td>
<td>.879</td>
<td>-5.004</td>
<td>.922</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sad</td>
<td>p9</td>
<td>-6.470</td>
<td>1.016</td>
<td>-5.045</td>
<td>.951</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p10</td>
<td>-4.926</td>
<td>.934</td>
<td>-3.791</td>
<td>.806</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Symbols</td>
<td>neutral</td>
<td>p9</td>
<td>-6.249</td>
<td>.960</td>
<td>-5.049</td>
<td>.968</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>happy</td>
<td>p9</td>
<td>-6.052</td>
<td>.906</td>
<td>-5.415</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sad</td>
<td>p9</td>
<td>-6.579</td>
<td>1.048</td>
<td>-5.057</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p10</td>
<td>-5.674</td>
<td>1.030</td>
<td>-3.847</td>
<td>.881</td>
</tr>
</tbody>
</table>
Table 9

<table>
<thead>
<tr>
<th>Img</th>
<th>Icons</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Mean</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutral</td>
<td>10.234</td>
<td>.769</td>
<td>10.814</td>
<td>.958</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>9.991</td>
<td>.799</td>
<td>10.382</td>
<td>.822</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>10.983</td>
<td>.974</td>
<td>11.440</td>
<td>.987</td>
</tr>
<tr>
<td>Cartoony</td>
<td>Symbols</td>
<td>Neutral</td>
<td>10.155</td>
<td>.941</td>
<td>11.271</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>9.979</td>
<td>.947</td>
<td>10.788</td>
<td>1.076</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>10.508</td>
<td>.890</td>
<td>11.150</td>
<td>1.019</td>
</tr>
<tr>
<td>Photo</td>
<td>Icons</td>
<td>Neutral</td>
<td>8.907</td>
<td>.777</td>
<td>7.842</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>9.310</td>
<td>.800</td>
<td>7.988</td>
<td>.805</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>9.010</td>
<td>.860</td>
<td>8.201</td>
<td>.769</td>
</tr>
<tr>
<td></td>
<td>Symbols</td>
<td>Neutral</td>
<td>9.089</td>
<td>.806</td>
<td>8.383</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>8.683</td>
<td>.815</td>
<td>8.068</td>
<td>.770</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>9.033</td>
<td>.925</td>
<td>8.395</td>
<td>.944</td>
</tr>
</tbody>
</table>
Chapter 6: General Discussion

My dissertation examined perceptual factors of cartoony faces, especially in comparison with photorealistic faces. In Chapter 2, I investigated a claim put forward by a popular, non-academic study of comic imagery: that with simpler faces, we are better able to project ourselves into the comic. I found no evidence supporting that hypothesis. Using the IAT to measure associations, I first found that participants did not associate photos of themselves with simple cartoony faces any more than complex cartoony faces. I then found that people associate photographs more than simple cartoony drawings they had made of themselves with self-words (e.g., “I”, “me”, “mine”, etc.). This suggests that we do not see ourselves in comics due to their simplicity. An alternative suggestion is that images with cartoony low-level features are easier to process, and that ease of processing may underlie their popularity.

In Chapter 3, I investigated whether manipulating the low-level features of face images to make them more cartoony would make it easier to discriminate expressions on those faces. I first found evidence that the more cartoony a face gets, the easier it is to discriminate the expression on it, and that this difference is greatest at very fast presentation times. EEG evidence further suggested that this difference was due to low-level features in the image. These results taken together add to a view of cartoony images serving to transmit information efficiently.

In Chapter 4, I tested whether there was evidence that ‘upfixes,’ symbols that occur above cartoony faces in comics, are easier to process as cartoony images rather than as photographs. Using two experiments to study how cartoony or photographic faces are understood in the presence of upfixes, this paper produced two notable findings. First, cartoony images demanded less overt visual attention, as measured by gaze deployment. This suggests that they were easier to process. I also found that, in conditions with ambiguous information, responses
indicated that cartoony faces were more congruent with upfixes than photorealistic faces. This may suggest that a screen which is entirely cartoony seems more congruent in meaning to participants because of its congruency in media style. These results support the findings from Chapter 3, showing that cartoony faces are easier to read than photographic faces, and this ease of information processing may underlie the use of symbols in comics.

Finally, in Chapter 5, I created cartoony and photographic faces that had mouth configurations that were not associated with recognizable facial expressions, and assigned them meaning for each participant. I found that faces with novel or non-canonical expressions did not evoke an attenuation of the N170 after a training session in which they were associated with a canonical/iconic emotional expression, whereas iconic stimuli did. I also discovered a different pattern of results before and after training for cartoony stimuli compared to photorealistic stimuli. Photorealistic stimuli evoked an attenuated N170 following training whereas the N170 elicited by cartoony images remained the same amplitude. Behavioral results further revealed that it was easier to learn to identify the emotion associated with novel expressions. Results from this series of studies all further point to an increasingly consistent story: that cartoony images are easier to process compared to more photorealistic representations. These convergent findings are consistent with the view that this ease of processing may be an important factor in the preferential use of cartoony images (e.g., in texting, in children’s media, in signage, etc.). Cartoony images may also add easily accessible information to more photorealistic images, as we saw in the case of upfixes.

In the next sections I discuss how my findings from this thesis fit into current understandings of how cartoony media differ from photographic media, and where each may be best used. There are several ways of looking at how cartoony faces are processed. There are the
proposals made by the artists who use them – for instance, explained by McCloud in his seminal book (McCloud, 1994), the perspective from research looking at face perception, and visual language theory. Each has a slightly different take on how cartoony faces fit into communication and their relationship to photographic faces.

Understanding comics

There have been attempts by artists to explore the idea of cartoony images and what differentiates them from photographs. Chief amongst these is Understanding Comics, which describes different types of cartoony images based on how abstract/concrete they are, how iconic/symbolic, and how simple/complex they are. In Chapter 2, I found results that suggested that there was no evidence supporting the idea that we see ourselves in simple cartoony faces, or associate their characters with ourselves. However, the results from my other studies also help shed light on some of the questions that McCloud and others have brought up.

There have been other attempts at categorizing symbols which appear in comics and animation (Walker, 1980), usually focusing on how symbolic content is clearly a defining trait of comic media. However, the idea that comics include symbols in a way that is unique to the form of the media (cartoony images) seems unlikely based on my research. While there are constraints about where symbols can be around a face, and what symbols are allowed (Cohn et al., 2016b), it now seems more likely that a comic image itself – apart from the imagery that comprises it, may have a structure that is culturally learned and that cartoony image is not strictly necessary for that structure to operate (further evidenced by the fact that the ‘linguistic’ structure of a comic can be different from culture to culture (Cohn, 2009). Instead, my research would predict that having a linguistic structure to a comic is capitalizing on the ease of processing that cartoony images can aid with.
One of the things that McCloud focused on most heavily when examining comic images was how different types of cartoony images could be divided, rated, and placed onto spectra, extending from the more iconic (or ‘realistic’) to the more symbolic (essential in ‘meaning’) and also more or less abstracted (see Figure 32). It is easy to see a Piercian, semiotic influence here – in that the images are being viewed, based on their form, as more like icons or symbols, and words are seen as an extension into symbolism while simple shapes are seen as an extension into...
abstraction, both different directions away from the photographic face. It is worth questioning, after the 8 experiments described in this thesis: are the distinctions described above meaningful?

Certainly, the idea that cartoony images always emphasize meaning based on their simplicity is not necessarily true – although the drawings they made of themselves were heavily simplified, the participants from my study in Chapter 2 were less likely to associate them with themselves than they were photorealistic images. However, the more efficient processing of cartoony faces compared to photorealistic faces revealed by the findings described in later chapters could suggest that cartoony-photorealism exists on a spectrum of efficient communication. The lack of N170 habituation to cartoony faces revealed by the data described in Chapter 5 may be because cartoony faces were seen to communicate expression without achieving ‘faceness.’ That is, photorealistic faces may engage multiple stages of processing each of which contribute to ERPs like the N170, and whereas cartoony faces may engage some of these stages (e.g., for individual features and the emotions they represent) they may not engage them all (e.g., the face as a holistic whole). This can lead to N170 results which appear similar to those evoked by photorealistic faces in some ways but not others – namely, a lack of habituation or inversion effects (Chammat et al., 2010; Sagiv & Bentin, 2001). Cartoony images and photorealism can be seen as opposing poles of a spectrum. An interesting outstanding question is whether there is there a certain level of cartooniness where a face begins to be processed more featurally than holistically, and therefore less like a face and more like another category of object.

Face perception

Much of this paper relied on interpreting face-sensitive ERPs, and in comparing different types of face stimuli. It is worth revisiting the face research which has supported the experiments
presented throughout this thesis and to ask if the results from these 8 studies have informed or
added to what we know about how faces are processed.

Prior to the research presented here, there have been few studies comparing photorealistic
faces to cartoony faces. What does exist suggests that cartoony faces are processed differently
than photorealistic faces. For instance, one study showed that emojis, one of the more popular
types of cartoony faces showed a different pattern of N170 results compared to photographic
faces – larger when upright, and reduced when inverted (Churches et al., 2014). Another study,
using schematic faces, found they evoked N170 magnitudes in the same range as photorealistic
faces when upright, but like the emoticons showed attenuation of the N170 for inverted faces
(Sagiv & Bentin, 2001). Studies have also found that cartoony faces are easier to process for
people on the autism spectrum (Rosset et al., 2008), which may be because cartoony faces tend
to be processed in a more featural manner compared to photographs (Prazak & Burgund, 2014).

Although these studies supported the idea that cartoony faces could be processed in a way
that is unlike photorealistic faces (i.e., more featurally, and without inversion effects), the issue is
complicated by the successful use of schematic faces in lieu of photorealistic faces in research
(Henderson et al., 2003; Leinbach & Fagot, 1993; Tomalski et al., 2009). Cartoony faces have
also shown N170 modulation by expression, with scrambled schematic faces showing a smaller
N170 than neutral faces, which was smaller than expressive faces (i.e., happy and angry
expressions (A. Maratos, Garner, M. Hogan, & Karl, 2015; Krombholz et al., 2007). These could
be compared to N170 results showing greater N170 amplitude for expressive faces over neutral
faces on photographs (Blau et al., 2007). These studies show that, in many contexts,
schematic/cartoony faces elicit the same experimental effects as photorealistic faces, and so
using them interchangeably was not unreasonable.
The data presented in this thesis can help resolve the question of whether cartoony faces are merely simpler photorealistic faces, and if there are qualities to cartoony faces which might help explain their widespread usage. In Chapter 3, I found that, while cartoony faces do indeed elicit neural signals which look ‘face-like’, as measured by ERPs, and that this is driven by low-level features, as would be expected as a ‘schematicized’ version of a photorealistic face, the degree to which the simplicity, exaggeration, and contrast affects the perception of an image has been generally under-estimated. While past studies had found that participants could discriminate fearful expressions on a photographic face above chance when presented at 33ms (Pessoa et al., 2005; Whalen et al., 1998), my research showed that for the most cartoony faces, all expressions were still almost at ceiling for discrimination at 17 ms (and indeed, at chance for photographs). This represents a vast difference between photographs and cartoony faces. In addition, and in support of the results described in Chapter 3, the upfix experiments described in Chapter 4 revealed that photorealistic depictions of faces evoke more fixations and dwell time than do cartoony faces. These results suggest that, while a cartoony face may be merely a simplified photorealistic face, this simplification can so exaggerate the results collected from participants that they could confound other effects a research may be looking for. In essence, at least in some contexts, they might as well be different types of stimuli.

The results described in Chapter 5 further call into question the idea that cartoony faces and photographic faces can be used interchangeably in psychological experiments. These findings showed that, while N170 habituation occurred for photorealistic faces, no such effect was shown for the cartoony faces that I used. This is significant because whether or not a stimulus is habituated to by the N170 has been used to discriminate between face and non-face stimuli in other studies (Maurer, 2008). This may suggest that the cartoony faces were simplified
past a point of being perceived as ‘real’ faces by the visual system, which is at least a gentle warning against using such faces as if they could be stand-ins for photorealistic faces within the context of psychological research.

Visual language

Finally, another branch of research that cartoony figure into heavily is in research based around the Visual language theory – that is, research looking at how comics are perceived, not just in their inclusion of cartoony faces but also in their syntactic form, visual vocabulary, and cross-cultural differences. Just as my research can inform how face stimuli are used in perception research, it also has implication for systems of iconography and how they are used in comic media and animation.

Most of the research done around visual language theory has examined how different aspects of comic media represent how cartoony images are processed on the syntactic level – for instance, how the framing of a comic image influences it (Cohn et al., 2005), or how the ordering of panels is constructed in a sentence-like structure (Cohn, 2014; Medley, 2010). However, increasingly there has been research into the smaller elements of cartoony media – the ‘semantics’ – operates. There is evidence that different cartoony images can have different meanings in different cultures (Cohn, 2007), and that there may be spatial relationships between symbols and icons (such as upfixes) in comic media, that are required for comprehension (Cohn & Murthy, 2013; Cohn et al., 2016b), as well as examinations on how other small lines or symbols that appear in comics are perceived (Cohn & Maher, 2015; Walker, 1980).

This paper adds to this research by showing that cartoony images may encourage participants to look for meaning between symbols and icons, such as how in Chapter 3, participants were more likely to rate emotionally ambiguous upfix/face dyads as matching if they were both cartoony than if the face was photographic. This supports the idea that there could be a
‘schema’ of how an upfix and face relate to each other that participants have learned from comic media. The ease of processing cartoony images, found in several chapters of this thesis, also help explain the presence of symbolic imagery in comics in general, and indeed the popularity of comic media and animation as a whole. That is, I have provided evidence that cartoony images are processed so easily in comparison to more photorealistic images that the presence of cartoony image dense media such as comics and graphic novels seems reasonable – whereas a photographic version would be much more onerous to process. Finally, the research from Chapter 5 also shows how easily cartoony images can be reassigned, and learned. Such an ability is crucial for underpinning any visual lexicon of symbols that people must be acculturated to through reading comics or watching animated movies.

Points of Consideration

In the following sections, I explore ideas that my thesis touched upon but did not fully explore. These ideas are worth examining as general points of consideration around the methods I have used in this thesis as well as my findings in each chapter.

Emojis

One of the most popular forms of cartoony faces is the emoji. However, the emoji is also one of the best examples of how seemingly iconic cartoony faces are often used symbolically. For instance, the crying tears of joy emoji is considered to be the mostly widely used emoji world-wide, and is influenced both by western and Japanese comics (Moschini, 2016); however, neither it nor any other emoji is agreed upon to have universal meaning, at least as measured by user understanding (X. Lu et al., 2016). It appears that it can be very difficult with emoji faces to determine what will be truly iconic and culturally universal, and what we only recognize symbolically (Miller et al., 2016; Miller, Kluver, Thebault-Spieker, Terveen, & Hecht, 2017). So while it may be easy to assume that emojis are used to replace non-verbal facial information
from conversation (McIntyre, 2016), the fact that their meanings vary so much from culture to culture suggests that emojis may represent an aspect of language unique to cartoony imagery that this paper was unable to approach. If future research were to look for an avenue where language and cartoony faces intersect, the emoji would be a likely candidate.

Research has revealed several things about the use of emojis, but none of it has definitely shown what an emoji is ‘for.’ For instance, as has long been intuited by their designers, emojis can clarify emotional sentiment. A recent study showed that a computer model made to gauge the positivity or negativity of an event based on tweets was able to do so with less variance when there were emojis involved (Ayvaz & Shiha, 2017). While an algorithm cannot be assumed to map onto human intention, it does mean that people appear to use certain emojis more often in positive contexts versus in negative contexts. Having emojis accompanying certain contexts appears hardly ground-breaking, and when described in a short manner, it seems clear what role emojis hold.

However, the difficulty is that tying emojis to emotional contexts (e.g., happy, sad, humorous) only works generally, and begins to fail when looking at specific instances. Emojis are also influenced by the context of the message and the sentences they are attached to – making machines worse at developing an algorithm to determine their meaning in any specific instance (Hakami, 2017). The meaning of emojis is also influenced by the graphical style platform they are presented on (Miller et al., 2017) (i.e., the same expression looks different on an android phone versus an iPhone versus on a computer, slightly altering intended meaning).

What makes emojis most notable, however, in the context of this thesis is that they are so heavily used for linguistic purposes. For example, emojis may be more likely to accompany specific words (Barbieri, Ronzano, & Saggion, 2016), but they can also be used alone, or as tone
enhancers for text, or, in rare cases, as a direct translation of text (Magnus, 2017). The use of emojis in so many linguistic roles is revealing – each of these uses is similar to how tone is added in informal language. But the use of emojis to affect linguistic tone is not necessarily what would have been predicted about them – it would imply that the emoji is being used not necessarily as a replacement for facial expressions, but rather as linguistic nuance. A further break down of emoji uses could contain other linguistic –type roles: softening an online message, reinforcing the meaning of a message, acting as an expression of the user’s unique personality, replacing text to provide meaning in a sentence, or altering the meaning of a text (as with a sarcastic tone) (McIntyre, 2016). And yet, the most obvious purpose of an emoji does have merit: the most commonly used types of emojis are iconic to human faces, suggesting they may, at least some of the time, be replacing in-person facial cues, simply with a different type of communication (Hakami, 2017; Ljubešić & Fišer, 2016; X. Lu et al., 2016; Magnus, 2017). It is also possible that the emoji is filling multiple roles: replacing facial information while also adding linguistic nuance into sentences which is otherwise difficult online.

So what do the results from this thesis add to this? As I stated above, emojis are perhaps the biggest target for trying to find a cartoony face which is being used both as an icon to resemble photorealistic faces while also filling communicative or even linguistic-type roles. My research from earlier chapters of this paper suggests a possibility for why: cartoony faces are not actually filling a special role that photorealistic faces could not – if you made very small photographic faces and inserted them into text to add emotional tone, it may indeed be successful. The difference that encourages the use of cartoony-type emojis may instead be partially explained by my results: that cartoony images require less time to process, and act as clearer stimuli which can be added to text without slowing comprehension down. For this to be
confirmed, however, further studies would need to be done, as it is possible that the insertion of cartoony faces into text in the manner of emojis has fundamentally changed how they are used and perceived compared to something such as comics.

Cartoony uses beyond media

Throughout this paper, I compared photorealistic faces with cartoony faces, looking for places where cartoony images would have a perceptual advantage. However, there are other areas of society where cartoony images are being favored over more photorealistic representations as well. For instance, there are innumerable examples of cartoony images being used for civic engineering projects, such as highways (Janda, Volk, & Lloyd, 1935; Shinar & Vogelzang, 2013) and subway maps (Castañeda, 2012), usually with the idea that cartoony imagery will be easier to spot and recognize (Garvey & Kuhn, 2011). Such examples can clearly be explained by extrapolating my results: if cartoony imagery is easier to discriminate expression information on due to low-level features, it is not surprising that cartoony imagery would also be easier to perceive on the street. However, one area that was outside of the scope of this paper is instances where cartoony images are used to communicate more complex types of information than expressions (or road signs). One of the most common examples of this is in medicine – cartoony images are often used in medicine to communicate medical information (e.g., symptoms) without text, making the information more culturally universal.

Medicine is also a good medium for understanding how cartoony images may support other types of information systems because there is much less room for error in creating effective symbols – if adoption of an approach to displaying medical information is going to become policy, it must show an improvement over established methods of how we represent medical information already. In fact, the American National Standards Institute requires that any
pictogram must be agreed upon by at least 85% of the people who will have to follow the pictorial instructions (Clawson et al., 2012; Mansoor & Dowse, 2004).

For example, in one study, researchers examined a system of pictograms intended to instruct people how to administer medicine. They compared different pictogram instructions amongst Xhosa (the language spoken by a South African ethnic group) speakers to see which were best understood by the populace. The researchers found that the best way to improve on symbols was to consult people within the target population, hold focus groups, and assess at each stage of improvement (Mansoor & Dowse, 2004). Thus, cartoony images here were used because they prevented the necessity of written language. The idea of creating universal cartoony image systems for understanding how to use and administer drugs seems to be catching on; the Spanish government has also created a legally binding set of universal pictograms for use in medical packaging – which is also based on their own governmental research (Fierro et al., 2013).

The findings from this thesis may have implications for understanding how to make effective cartoony symbols for such systems – for instance, relying on iconicity, or instituting brief training systems based on how easily new information can be tied to cartoony images, or by simplifying cartoony images as much as possible to create the easiest possible processing of new cartoony images. However, as the research presented in this thesis was limited mostly to faces, the degree to which it applies to pictorial systems is limited. One of the lessons from the upfix studies in Chapter 3, however, is to encourage exploring stimuli based on existing conventions that people already understand. While the cartoony upfixes I used did demand less attention than photorealistic upfixes, the attention given to any upfix paled compared to the face stimulus. It would be worth advising others that, should you create pictograms containing faces and symbols,
the faces will always be the most attended to part of the image – which could change what is learned and how.

Future Directions
Faces are open books

It may be best to think of the cartoony image as simply a very attractive, or efficient, medium. However, is there a type of information that cartoony images excel at carrying? This is a more difficult question to answer. It certainly appears that many of the common and naïve uses of the cartoony image aim to express emotion or some other internal state. For instance, that is often what is being portrayed in children’s use of cartoony imagery (Misailidi, Bonoti, & Savva, 2012; Toshijuki, Takasaki & Mori, 2007), the most common use for emojis (Ljubešić & Fišer, 2016; Miller et al., 2017; Moschini, 2016), and expressing how someone feels inside, may underlie why cartoony picture systems work well for patients to describe their symptoms to a doctor (Ferreira-Valente, Pais-Ribeiro, & Jensen, 2011; Tack et al., 2014).

However there are also examples to contradict the idea of cartoony images only excelling with expressing internal states of a person – that concept cartoons and other scientific cartoony images are so successful at imparting scientific information (Keogh, 1999; Lin, Lin, Lee, & Yore, 2015; Naylor & Keogh, 2013) seems to suggest that other types of information aside from emotions are just as readily transmitted through cartoons. Certainly, the symbolic system for labelling stations on the Mexican metro stations involves no expression of internal states by anyone, yet remains popular (Brenišínová, 1973) Perhaps the fact that longer comics may show less of an improvement over bare text (Jee & Anggoro, 2012; Tatalovic, 2009), as well as the fact that concept cartoons seem to work better when carefully constructed with both images and text (Naylor & Keogh, 2013) may be revealing.
It seems likely that where the cartoony image naturally excels is in transmitting simple information including scientific concepts that may demand textual support, conveying emotional states, setting conversational tone, and attributing value (e.g. the feeling of finding something cute or innocent). In addition, studies looking to use cartoony icons and symbols in medicine also seem to support the idea that simple information is what a cartoony image is best at transmitting – more complex information requires a more careful tuning of cartoony imagery to be understood (Green & Myers, 2010; Mansoor & Dowse, 2004).

The question of what kinds of information a cartoony image is best at representing is clearly understudied. The experiments in this thesis have unfortunately used only facial expressions on different types of faces, and so it not clear if the results would generalize to different types of cartoon and photographic images. While it is reasonable to assume – based on their widespread usage for better traffic signage – that cartoony images are generally processed more easily, even outside of faces, but what is less clear is if cartoons would hold their advantage for other aspects of images. For instance, would cartoony faces be superior for facial identity recognition? Discriminating facial identity requires the integration of many types of features, e.g., eye shape, width, eyebrows, colour, etc. The simplicity of cartoony faces may result in crucial information being lost – although, depending on the cartoony face, it is clear that certain features can also be emphasized in cartoony faces, leading to a caricature advantage in some circumstances (Mauro & Kubovy, 1992). Similarly, while cartoony images are processed more featurally (Sagiv & Bentin, 2001), making them easier for people who have a deficit in holistic face processing to respond to (Joseph & Tanaka, 2003), responses to photographs may show a processing advantage in tasks requiring holistic face processing.
Cultural understanding of visual language

Just as emojis can be interpreted differently from place to place, there is strong evidence that cartoony images can be interpreted differently in general from culture to culture (Cohn, 2009), including images which may be drawn from a long cultural tradition (Clark, 2009). Thus, in the context of comics and animation, it is reasonable to expect that there are more layers of cultural information than may be immediately obvious. The studies I performed in this paper focused on situations where the level of symbolic information was still relatively low. A study that used stimuli with multiple layers of symbolic meaning, such as in comics that can have different frame types, symbols around characters, and symbols on characters may better reveal how symbols are approached in complex stimuli.

Having a good cultural understanding of cartoony images is becoming increasingly important. Cartoony images have begun to be used in areas where cultural specificity may create roadblocks. For instance, cartoony imagery is used online and in VR environments for the creation of avatars (Jack, Garrod, Yu, Caldara, & Schyns, 2012; J.-E. R. Lee et al., 2014), and if art style of the cartoony images influences how people see themselves and others in their environment, it becomes an important hurdle for creating virtual spaces that are universally appealing. Another unexpected area where cultural aspects of cartoon imagery could appear is in the creation of robots. Robots are increasingly being created with simplified, animated, emoji-type faces (Armus, 2015; Blow, Dautenhahn, Appleby, Nehaniv, & Lee, 2006; Hoa & Cabibihan, 2012; Young, Xin, & Sharlin, 2007), however as mentioned above, emojis are anything but universal, and many such cartoony faces rely on symbolic lexicons taken from comics and animations from individual countries which are also not in any way universal.

Creating a set of cartoony faces and symbols which can be universally used and interpreted is a challenge which research could inform on. While in this thesis I have shown that
cartoony faces have many advantages compared to photorealistic images, the way to best construct a cartoony image so that it appeals to users and can be used best for communication is an area of study that needs a great deal more attention.

Conclusion

What perceptual or communicative qualities separate cartoony faces and photographic faces? That was the question that motivated the research summarized in this thesis. After 8 experiments, the most likely answer would be: cartoony images show processing advantages due to their low-level features, and possible additionally because of linguistic-type schemas we have learned from comics and animations. These advantages allow cartoony images to be used more effectively than photographs in contexts where rapid processing is important. This is a very simple answer for what turns out to be a very complex issue, of course. While cartoony images are indeed processed much more quickly than photographs, I only tested this hypothesis thoroughly with expression information. There could be caveats still to be discovered based on differences between schematic stimuli and fully cartoony stimuli as well as cultural biases in understanding different cartoony symbols.

Ultimately, the message that comes from these studies is not necessarily that we have been thinking of cartoony faces incorrectly; research premised on the assumption that they can function as be simpler photographic faces was supported, as was the idea that they are used in linguistic media, as was the idea that the low-level features of cartoony images is a central aspect that separates them from photorealistic depictions. What was missed by previous work is, I believe, the scale of the difference. Cartoony images in my experiments have been shown time and time again to be processed more easily than photographs, to demand less attention, to be discriminated more quickly, and to be learned faster. The gap between how easy it can be to respond to a cartoony stimulus and how difficult for a photograph cannot be overlooked – even if
a cartoony face and a photorealistic face are essentially two versions of the same type of stimulus, the cartoony face can be so different in how people respond to it that they should not be used interchangeably.

Moving into the future, this research should be taken as not only a curiosity, or even an admonishment, but hopefully as a signpost to the future, as cartoony images only become more and more used, more and more consumed, and more and more popular, following the explosive growth of new technologies.
Works Cited


https://doi.org/10.1016/j.neuropsychologia.2008.11.016
Cohn, N., Murthy, B., & Foulsham, T. (2016b). Meaning above the head: combinatorial constraints on


Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal Differences between Object and Face


https://doi.org/10.1080/21548455.2014.941040

https://doi.org/10.1016/j.neulet.2010.08.059


https://doi.org/10.3389/fnhum.2014.00213


https://doi.org/10.1007/s12264-013-1401-4

https://doi.org/10.1016/0013-4694(90)90139-B


https://doi.org/10.1145/769759.769774

https://doi.org/10.3389/neuro.09.018.2008

https://doi.org/10.3758/BF03210927

https://doi.org/10.1109/TPC.1998.661632

https://doi.org/10.3758/BF03200730

McIntyre, E. S. (2016). From Cave Paintings To Shakespeare and Back Again: What Are Emoji and Should I Be Afraid?, (May).


Rojas, S. L., Kirschenmann, U., & Wolpers, M. (2012). We have no feelings, we have emoticons ;-). In Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies, ICALT 2012 (pp. 642–646). https://doi.org/10.1109/ICALT.2012.180


Takasaki, T., & Mori, Y. (2007). Design and development of a pictogram communication system for


strongly influence responses in the fusiform face area. *Cerebral Cortex, 21*(1), 35–47. 
https://doi.org/10.1093/cercor/bhq050