5-2019

Familiarization through Ambient Images Alone

Nia Imani Gipson
University of Arkansas, Fayetteville

Familiarization through Ambient Images Alone

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Psychology

by

Nia Gipson
McDaniel College
Bachelor of Arts in Psychology, 2016

May 2019
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

____
James Michael Lampinen, Ph.D.
Thesis Director

____
William Levine, Ph.D.
Committee Member

____
Darya Zabelina, Ph.D.
Committee Member

Abstract

The term “ambient images” has begun to show up in much of the current literature on facial recognition. Ambient images refer to naturally occurring views of a face that captures the idiosyncratic ways in which a target face may vary (Ritchie & Burton, 2017). Much of the literature on ambient images have concluded that exposing people to ambient images of a target face can lead to improved facial recognition for that target face. Some studies have even suggested that familiarity is the result of increased exposure to ambient images of a target face (Burton, Kramer, Ritchie, & Jenkins, 2016). The current study extended the literature on ambient images. Using the face sorting paradigm from Jenkins, White, Van Montfort, and Burton (2011), the current study served three purposes. First, this study captured whether there was an incremental benefit in showing ambient images. Particularly, we observed whether performance improved as participants were shown a low, medium, or high number of ambient images. Next, this study attempted to provide a strong enough manipulation that participant would be able to perform the face sorting task perfectly, after being exposed to a high number (45 total) of ambient images. Lastly, this study introduced time data as a measure of face familiarity. The results found support for one aim of this study and partial support for another aim of this study. Time data were found to be an effective quantitative measure of familiarity. Also, there was some evidence of an incremental benefit of ambient images, but that benefit disappeared after viewing around 15 unique exemplar presentations of a novel identity’s face. Lastly, exposing participants to 45 ambient images alone did not cause them to reach perfect performance. The paper concludes with a discussion on the need to extend past ambient images to understand how to best mimic natural familiarity in a lab setting.

Table of Contents

Introduction

In the recent years, the literature on face recognition has focused on the role of ambient images. In this context, ambient images are defined as naturally occurring images of an individual's face, which capture the idiosyncratic ways in which a face can look different (Ritchie & Burton, 2017). Through exposure to ambient images, participants can improve recognition of unfamiliar faces (Andrews, Jenkins, Cursiter & Burton, 2015; Dowsett, Sandford & Burton, 2016; Kramer, Ritchie & Burton, 2015; Murphy, Ipser, Gaigg & Cook, 2015; Ritchie & Burton, 2017; Sweeney & Lampinen, 2012). Recent literature is developed around the theory that repeated exposure to novel ambient images of a face allows us to form a stable representation of an identity, which will aid with later facial recognition. It is believed that the use of ambient images in studies mimics how we naturally become familiar with faces (Burton, Kramer, Ritchie, & Jenkins, 2016).

The importance of ambient images in face recognition first became apparent in a study by Jenkins, White, Van Montfort, and Burton (2011). Participants in this study were assigned with the task of sorting face images into piles by identity. Specifically, they were given 40 images of two different unfamiliar identities. Yet, when completing this task, on average participants generated 7.5 piles, thus overestimating the number of identities present. The authors concluded that the participants had difficulty with this task because they failed to discern slightly varied images of an individual as still depicting the same person.

Jenkins et al. (2011) were not the first or last to demonstrate poor performance with judging two images of an unfamiliar person as being the same person. Matching tasks are a very popular paradigm, which directly tests this ability. These tasks require participants to judge whether two face images, presented side by side, depict the same identity. Even though the images are presented simultaneously, participants still struggle with this matching task (Bruce et al., 1999; Bruce, Henderson, Newman & Burton, 2001; Kemp, Towell, & Pike, 1997; Megreya & Burton, 2008; Megreya, Sandford & Burton, 2013; Kramer & Ritchie, 2016; Papesh, 2018). These studies show that slight variations between photos of a single identity are not well tolerated when viewing unfamiliar faces. An earlier study found that differences in facial images as minuscule as change in angle could significantly decrease performance when matching a target image to its related image within an array (Bruce et al., 1999). A more recent study manipulated whether both images featured the target with glasses, no glasses, or both glasses and no glasses. The results show that performance drops when the presence of glasses are not kept consistent between photos (Kramer & Ritchie, 2016). Additionally, another study introduced the difficulty of matching a live person to a photo. In this task, a target was directed to stand in front of room adjacent to a lineup array. Yet participants only choose the target out of the lineup correctly about 67% of the time (Megreya & Burton, 2008). There is only a select group of individuals called super-recognizers who consistently perform better than average on face matching tasks. One recent study found that when given the Glasgow Face Matching Task (GFMT), a common face matching paradigm, super-recognizers are accurate on 97% of trials, while matched controls were only accurate on 87% of trials (Bobak, Dowsett, & Bate, 2016).

Our difficulty with matching images of unfamiliar persons has strong implications to real-world scenarios. A study by Kemp et al. (1997) recruited participants who worked in a supermarket and who were often tasked with matching costumers to their photo ID. For this study, the cashiers were required to make judgments on whether an ID card matched the identity of the volunteer. Half of the volunteers presented a true match ID, while the remaining volunteers presented a non-matching ID. Despite their presumed expertise with face matching, these cashiers were only correct on 67.4% of the trials. Similarly, a more recent study had bank tellers, notaries, and students complete a face-matching task with photo IDs. The results indicated that neither bank tellers nor notaries were any better than undergraduate students at successfully matching a person to an ID (Papesh, 2018). Again, when similar methods were used with passport agents, researchers found no difference between these experts and undergraduate students (White, Kemp, Jenkins, Matheson & Burton, 2014). The only study that could find evidence for expertise was conducted with forensic facial identification examiners (White, Phillips, Hahn, Hill, & O'Toole, 2015). These experts were less impacted by face inversion tasks, which require a participant to make judgments on upside-down face images. However, these experts still performed equally to the control group when completing the GFMT. Considering these results together, these experts may have only experienced heighted ability in certain aspects of face perception, rather than an overall benefit. It may be necessary to disentangle these conflicting results before concluding that these experts are better at facial recognition than the general population. Despite this, the literature suggests that expertise cannot protect against natural deficits on unfamiliar face matching tasks.

It is important to note that in all the studies mentioned above, participants were asked to match images that were taken within the same day. A study by Megreya et al. (2013) found evidence that the ability to match faces gets worse when photos are taken as little as a month apart. Performance dropped from an 87.5% hit rate when matching photos taken minutes apart to a 67.5% hit rate when matching photos taken a month apart. Often experts are tasked with matching face images that were taken days, weeks, months or even years apart. For example, when matching ID photos, the image may depict the target when they were several years younger. If these results extend to experts, face-matching ability may be worse in real world settings than what has previously been described in the literature.

Unfamiliar Versus Familiar Face Perception

Many researchers have begun to refer to this shortfall in unfamiliar facial recognition as an inability to "tell face together" (Andrews et al., 2015; Jenkins et al., 2011). The literature shows a consistent trend of the difficulty of assigning the same identity to different images of an unfamiliar face. However, these same deficits are not observed when participants are given familiar face stimuli. The Jenkins et al. (2011) study mentioned earlier presented a clear demonstration of participant's tendency to consider variant images of an unfamiliar face as two separate identities. Though when the researchers gave the same tasks to participants who were familiar with the face stimuli, participants typically made no errors. This study highlights the clear distinction between familiar and unfamiliar face recognition. A sign of familiarity is the ability to accurately group images of an individual's face together despite how different the images are from one another. For familiar face recognition, within person variability is more tolerated than for unfamiliar face recognition.

Bruce et al. (2001) made a direct comparison of face matching performance with familiar versus unfamiliar face stimuli. They compared clear images to images of a person taken from a CCTV video recording, which were often poorer in quality. The results show that familiar faces were matched with 93% accuracy while accuracy for unfamiliar faces dropped to about 75%. These results support a larger idea that familiar face matching is more resistant to image variability than unfamiliar face matching. Participants are better able to recognize that two highly variant instances of a familiar face depict the same person.

In line with this idea, a few studies have also found that people tend to assign higher likeness scores to ambient images of someone they are familiar with. This result was first found in Jenkins et al. (2011). During this study participants were asked to provide likeness ratings, in which they judged the extent to which a face image of a celebrity resembled that celebrity. All celebrities used within this study were well-known, though participants were instructed to skip a celebrity if they were unfamiliar with them. In short, the authors quantified the popularity of a celebrity by how many participants indicated that they knew the celebrities well enough to perform the task. The results showed that celebrities that were more known among all participants also received greater overall likeness scores for all their ambient images. The authors conclude that as a person becomes more familiar, more images of them serve as good representations of how they look despite how variable these images appear. A more recent study was able to replicate these results and demonstrate evidence of a within-person correlation between familiarity and likeness (Ritchie, Kramer & Burton, 2018). Particularly, this study totaled the number of movies a participant saw of a celebrity and then assessed the likeness scores the participant gave different ambient images of that celebrity. The results indicated that the more movies a participant saw of that celebrity, the higher the mean likeness score they assigned to all images of that celebrity. Again, this concept of likeness is a relevant measure when it comes to "telling faces together." If a participant can judge more images as being a good representation of a target identity, we can expect that even extremely variant face images of the individual will still be judged to have some likeness to the target identity. It would be rarer for that participant to erroneously reject images of that identity despite how varied they may look if the participant is more willing to judge any ambient image as having good likeness. We might then expect that participants may not be as willing to rate images as having "good-likeness" when working with an unfamiliar identity. Alternatively, participants may only assign high likeness scores to images of unfamiliar faces that best match whichever images they originally saw of that unfamiliar face. In other words, if we have a broader cognitive representation of an identity, we likely saw many exemplars of that identity. Thus, there is an increase likelihood that any novel images of that identity will be match our representation of that face.

Other literature has also revealed clues on how unfamiliar faces are processed differently than familiar faces. Using a face matching task, Megreya and Burton (2006) attempted to determine if performance on a familiar face matching task could predict performance on an unfamiliar face matching task. The results indicated that unfamiliar face matching performance predicts familiar face performance only when the familiar faces are inverted. However, there is no correlation between familiar face performance and inverted familiar face performance on this task. Inverting faces has commonly been used as a method to disrupt configural face processing, thus making it harder to recognize a face image or derive identity from it (Diamond & Carey, 1986; Yin, 1969). According to the configural processing theory, we can recognize a face based on the relation of facial features to one another (Carey & Diamond, 1977). This results potentially suggests that like familiar inverted faces, upright unfamiliar faces may invoke a similar lower level of face processing (Megreya & Burton, 2006). By disrupting configural processing, face recognition is no longer automatic for participants. So, they must find other means of matching faces, such as matching the individual features of a face. In the case of unfamiliar face matching tasks, participants may also be required to match images using features instead of being able to rely on some stable pattern of the relationship of facial features. Configural processing might explain much of the variance we see in disparities between familiar and unfamiliar face matching tasks.

A more recent study found that participants were faster at distinguishing between unfamiliar and familiar faces than distinguishing between only familiar faces (Besson et al., 2017). In this study, participants were shown a series of familiar or unfamiliar faces and asked to make a judgement on each face. When shown only familiar faces, participants were asked to make an identity judgement (i.e. "Is this a picture of '_____'"). When shown a mix of familiar and unfamiliar faces, participants were asked to make a familiarity judgement (i.e. "is this a picture of a familiar person"). Even when the faces were inverted, reaction times were slowest when differentiating a target familiar face among many other familiar faces, then when differentiating familiar faces among many unfamiliar faces. Again, these results suggest that unfamiliar faces are processed differently from familiar faces. The faster reaction time when differentiating familiar faces from unfamiliar faces might indicate that familiar face stimuli are categorized separately and are distinct from unfamiliar face stimuli.

Multiple Image Exposure to Familiarity

Knowing that there is evidence of clear differences between familiar and unfamiliar faces, a next area of research to explore is how an unfamiliar face becomes familiar. As mentioned earlier, exposure to ambient images is one of the current theories on how an unfamiliar face can become familiar. Many studies have shown that by viewing multiple ambient images, participants can create a stable representation of a face (Bruce, 1994; Kramer et al., 2015; Burton et al., 2016; Ritchie & Burton, 2017). This can occur because when viewing different images of an individual's face we learn to focus on the traits that are most constant from image to image. Typically, it is the internal traits that are considered most stable (Ellis, Shepherd, & Davies, 1979; Robins, Susilo, Ritchie & Devue, 2018; Young, Hay, McWeeny, Flude, & Ellis, 1985). Internal traits refer to the spatial distance of facial features, as well as, the general shape and size of one's facial features. Such traits are not as variant from image to image, when viewing novel faces. A recent study by Robins et al. (2018) found that by editing images to only include the internal traits, participants were able to improve matching accuracy to other similarly cropped photos. The authors conclude that by having participant view only internal features of the face, participants could avoid focusing on transient features, such as hair, which can easily change from image to image. Instead these participants could focus on more stable internal features of the face.

Further evidence supporting the creation of stable face images comes from literature on face averages. A few studies have demonstrated that participants create a mean image of face images when shown multiple at once (de Fockert & Wolfenstein, 2009; Kramer et al., 2015; Neumann, Schweinberger & Burton, 2013). Many of these studies have participants view an array of about four photos for a few seconds. Immediately after participants are presented with either a morphed average of all four exemplars, one of the four exemplars, an image of a new person, or a morphed average of four new people. Participant are tasked with identifying whether these image types appeared within the original array. Consistently, the results show that participants are just as likely to choose the morphed average of the four studied images, as they are to choose a single exemplar of a studied identity. These results have been demonstrated when showing different celebrity faces at once as well despite the potential for a participant to use verbatim recall to remember the exact identities they saw (Neumann et al., 2013). In addition, studies have replicated these findings with unfamiliar faces (de Fockert & Wolfenstein, 2009; Kramer et al., 2015). One study presented a four-image study array of ambient images of an individual (Kramer et al., 2015). Again, participants were just as likely to indicate that an average of those four images were present, as to recall a single image that was displayed. This phenomenon is referred to as ensemble encoding and suggest that it might be natural to create an average when confronted with several related images (de Fockert & Wolfenstein, 2009). However, because this effect is seen not only when viewing images of the same person but as well with images of different people, it may not fully explain how we become familiar with faces. This phenomenon seems to occur too frequently to solely benefit the creation a stable face representation for identities. Also, in real life we may not be confronted with several different images of an individual at once. Rather we must be able to create a stable image of an individual even when views of a person are presented days apart.

An additional theory on how we represent faces was proposed by Valentine (1991). Under this model, each time a person is presented with a new identity, the face of that identity gets added to a "face space." The dimensions of this face space can be divided based on any characteristics that might help to set certain individuals apart (i.e. skin color, gender, familiarity, etc.). Identities that fall more central in the face space are believed to be more prototypical, while faces that are more distant from the center of face space are considered more distinct (Valentine, 1991). Each identity has an "attractor field," which determines how similar a face image must be to be grouped with that identity. The larger an attractor field, the more likely highly variant images of that identity might be accepted (Laurence, Zhou, & Mondloch, 2016). We can infer that familiarity or increasing the number of exemplar images that match an identity, would allow an attractor field to grow. However, this growth would only affect that single identity, which is why we do not typically see carryover effects of learning many images of one identity (Dowsett et al., 2016; Kemp et al., 1997; Papesh, 2018; White et al., 2014).

Whether we are truly forming a prototype face image of each identity we encounter, there is still abundant evidence that seeing multiple images can lead to improved recognition of unfamiliar faces (Andrews, et al., 2015; Dowsett et al., 2016; Menon, White & Kemp, 2015; Murphy et al., 2015; Ritchie & Burton, 2017). Many of these studies reveal that recognition performance improves as the number of exemplar images increase. Dowsett et al. (2016) had participants identify a target identity from a pile of 30 filler identities. After each trial, a new image of the target identity was given to the participant to study. This was done up until the participants had six target images in front of them when choosing. The authors found that performance improved each time the participant was given another exemplar image to study. Similar findings extend to recognition tasks as well. Murphy et al. (2015) had participants view either six images repeatedly or 96 unique images of each identity. The results show participants were significantly better at guessing how many identities were present when implicitly exposed to many exemplars (96 total) than when shown the same few images (six total) repeatedly. The authors highlight that exposure to variation is more necessary for facial recognition than repeated exposure.

Several designs further suggest that only exposure is needed to see these benefits, even if participants are not explicitly told that the images depict the same individual. During the same Murphey et al. (2015) study participants were simply shown images of faces but were not given feedback on how to properly group them. Specifically, participants were shown an array of 48 images and told to guess how many identities were present. There were always six images of each identity within the array, for a total of eight identities present. In one condition, the images were simply shuffled and replaced back in the array after each trial, so only six images of each identity were used throughout the whole study. In the other condition, the images were replaced during each trial, until participants were exposed to a total of 96 images of each identity. Again, the authors found that participants could more quickly correct their guess for the numbers of identities present when exposed to 96 exemplars rather than the same six. On a follow-up recognition task, participants were shown novel images of the target identities. Participant in the 96-image condition gave accurate judgements 80.8% of the time while those in the six-image condition gave accurate judgements 72.9% of the time.

In addition, a study published the same year utilized the sorting task paradigm from Jenkins et al. (2011) as a study phase found similar results. Participant in this study were either explicitly told or remained unaware of the number of identities present in the pile of face images (Andrews et al., 2015). These participants then completed a follow-up face matching task with the identities they sorted. There was no difference in performance on the follow-up face matching task, despite that participants in the informed condition created more accurate piles at study than those in the uninformed condition. Accurately identifying the identities did not lead to any added benefit than just simply being exposed to images of the identities. Both studies highlight that simple exposure to ambient images is sufficient in improving face recognition.

Several studies have gone further to show clear differences in performance when shown images that are highly variant versus images that have low-variability. Menon et al. (2015) provided participants with either a high-variant pair or low-variant pair of face images. This study used a face-matching paradigm, in which participants were asked if the high-variant or low-variant pair matched a novel face image. The results show that participant performed better with matching the test image when they were exposed to the highly variant pair. A more recent study used similar methods but employed an old-new recognition task. In Ritchie and Burton (2017), participants were given 10 high-variability images or 10 low-variability images at study. The low-variability images were all screen shots from a video interview, while the high-variability images were ambient images, taken at different times and locations. The authors found similar results, in which participants performed better when exposed to highly variant images.

Though performance is shown to improve when increasing number of exposures and increasing variability of the study images, there has yet to be any evidence that such exposure can lead a participant to perform as well with unfamiliar (or recently familiarized) faces as with familiar faces. A next direction to explore is determining whether simply exposing participants to multiple varying images of an unfamiliar face, can lead to similar performance as expected with familiar faces. As mentioned earlier, Jenkins et al. (2011) shows a clear distinction between performance with familiar faces and unfamiliar faces. When sorting celebrity faces performance was near perfect across participants; however, when sorting unfamiliar faces participants consistently created more piles than there were identities. This sorting task has been utilized in several other studies and similar results have been found. Participants consistently create more groups than there are identities, when they are unfamiliar with the identities present. Also, participants rarely mistakenly group two different identities into one pile, which is commonly termed a misallocation error. Instead, the most common error is that participants create several piles for one identity (Andrews et al., 2015; Balas & Saville, 2017; Jenkins et al., 2011; Laurence et al., 2016; Redfern & Benton, 2017b).

Other Factors Affecting Performance on Face Matching Tasks

As we continue to explore how unfamiliar faces might become familiar, it is important to note that other factors might affect performance on face matching tasks. Some previous studies have examined the role of factors, such as expressions and expectations, on performance during face matching/grouping tasks. For example, more expressive face images can have strong effects that either aids or harms face recognition (Mileva & Burton, 2018; Redfern & Benton, 2017a;

Redfern & Benton, 2017b; Redfern & Benton, 2018). In addition, depending on the directions given at test, participants can create their own expectations about the task and therefore perform accordingly. Both these factors are important to consider as research on facial recognition continues. Understanding the nature of the face stimuli that participants are exposed to and methods that might affect their expectations can allow researchers to better interpret results.

The literature on expressions and face recognition shows vastly different results depending on the task. One name-verification task required participants to accurately identify targets as either one of two identities (Redfern & Benton, 2018). The identities used were celebrities; however, none of the participants were familiar with them. They first received training to learn the difference between the two identities. Participants were either trained with highly expressive faces or faces low in expression. The results revealed that those trained with low expression images had significantly longer reaction times to high expressive faces than low expressive images. However, those trained with high expressive images performed equally at test with both low and high expression test stimuli. This study suggests that studying expressive faces still allows participants to build a representation of a face strong enough to extend to less expressive instances of that face. The opposite of this is not true, as participants who viewed only low expression images had slowed performance when shown high expression images. Perhaps there is an added value in learning a range of expression when learning a new face. This relates to earlier work that found that high-variability in general is beneficial to face recognition (Ritchie & Burton, 2017). Recognizing a face can be a slower process if an individual only has access to more neutral instances of the face. In addition, another study demonstrated improved performance when participants were shown faces with open-mouth smiles on a matching task, rather than neutral expressions or a closed-mouth smile (Mileva & Burton, 2018). The authors believe that open-mouth smiles allowed for more unique aspects of a face to become apparent, making it easier for participants to match instances of a face. In both studies, there appears to be some added information that unfamiliar viewers can obtain from more expressive faces that are not easily obtained when viewing less expressive faces.

Despite these findings, Redfern and Benton (2017b) did not find similar patterns when using the face sorting task paradigm. In this study, participants sorted images of neutral and expressive faces. When sorting expressive faces, participant created significantly more piles and had significantly more misallocation errors (grouping two separate identities together). The results suggest that participants were more likely to interpret face images high in expression as being two separate people. Another study by Redfern and Benton (2017a) conducted that same year found slowed reaction times, when participants were tasked with identifying a face high in expression than a face low in expression. These results suggest that expressions may simple be useful during learning but harmful when discerning instances of an unfamiliar identity.

In addition, based on previous research, it is difficult for participants to reconcile two variant face images of an unfamiliar identity as depicting the same person. We can expect that the condition with more expressions would also have more images that are highly variant, thus making this sorting more difficult. Perhaps the demands of the face sorting task prevented any benefits of face expression. In the name-verification task and the matching task, participants were only presented with one or two images at a time. This allowed them to allocate more attention to the images presented and perhaps made it easier for them to notice relevant traits that could help them perform the task. However, during the sorting task participants have 40 images that they must make decisions on and it is up to them to choose how to allocate their attention during this task. Perhaps to deal with the demands of the task, participants might have allocated less attention to each image and therefore might have missed out on relevant traits that would have allowed them to be more successful on this task. This interpretation leads us to the importance of expectation. If we can get a participant to properly allocate their attention during this sorting task, their performance can greatly improve.

A study by Andrews et al. (2015) highlighted the importance of participants expectations on face recognition tasks. In this study, participants were told that only two identities were present in the sorting tasks. As a result, they accurately sorted the identities into two piles (Andrews et al., 2015). This study provides evidence that unfamiliar face performance on this task can mimic familiar face performance not only due to the manipulation but also due to the participant's expectations. Again, this result might have occurred because the researchers were better able to focus the attention of the participants. Additional data on completion time shows that participants took just as long to complete the informed sorting task as the uninformed sorting task. This data further supports that participants did not simply take more time to improve performance but were able to allocate cognitive resources to most successfully complete this task. It is also important to note that during this task, participants were able to use an additional cue to make judgements. Whenever a participant was unsure of the identity of an image, they could find the pile that provided the best match, since they knew the image belonged to one of two identities. This cue would not have been available to those in the uninformed sorting task.

Based on these results, it will be important to control for participant expectancy during the current study. We want to observe changes in familiarity separate from any changes in expectation. Additional measures, such as completion time, might also help to more accurately assess the difference between performance with familiar and unfamiliar faces. Again, the Andrews et al. (2015) was the first study to add time data and demonstrated that regardless of the participants' expectation they still took roughly seven and a half minutes to complete an unfamiliar face sorting task. The current study will determine whether there are time differences when performing a familiarized versus unfamiliar face sorting task. If we measure time during this sorting task, we might be able to uncover further distinctions between performance with unfamiliar and familiar faces.

Study Overview

The current study aimed to demonstrate whether viewing ambient images alone can cause an unfamiliar face to become truly familiar. We measured familiarity using the face sorting task from Jenkins et al. (2011). Participants studied identities of varying set sizes to determine if ambient image exposure is all that is needed to make unfamiliar faces performance equal to familiar face performance. Both time and accuracy were recorded during this task to assess whether performance with recently familiarized faces is equal to performance with unfamiliar faces. Perfect performance on this task would indicate high familiarity with the recently studied identity.

Previous studies have manipulated the participant's familiarity with novel faces by varying exemplar set sizes. However, these studies typically only used a low versus high familiarity groups (Bruce et al., 2001; Menon et al., 2015; Murphy et al., 2015). The current study attempted to extend the literature by including a three-level familiarity manipulation (low, medium, high) to better determine if there is indeed an incremental benefit in showing more ambient images. Andrews et al. (2015) found that whether a participant sorted face images into two piles of 20 or several smaller piles, they still had equal performance on a follow-up face matching task with the same identities. Performance also did not differ in regard to completion time, as both groups took roughly seven minutes and 30 seconds to complete the unfamiliar face sorting task. This finding informed two major questions for the current study. First, we further investigated the usefulness of time data on this task. Though Andrews et al. (2015) did not find any difference in completion time when comparing two unfamiliar sorting tasks, there may be some differences in time data when comparing an unfamiliar sorting task to a familiar sorting task. Second, the authors made an important interpretation of these results, in which they suggested that this finding shows there is no added benefit after a certain amount of ambient image exposures. In short, the few images that participants could group together in the uninformed task provided enough information for them to accurately match photos in the next task. In addition, in line with literature on stable face representation, Jenkins and Burton (2011) used a computer program to create stable face images. They found that after combining 12 images of a single identity's face, the average became very stable and unaffected when adding additional images. These accounts both suggest that there may be some limit to the benefit of ambient images after a certain number of exemplars. By using three levels of study, varying in number of images presented, we aimed to find evidence for or against the incremental benefit of ambient images.

We first hypothesized that viewing many (45 total) ambient images would allow participants to accurately sort the images into two piles, thus indicating that the participant was familiar with studied identities. Second, we expected to see incremental improvements in performance as participants were exposed to more and more ambient images. In other words, performance would steadily improve from the low image exposure condition to the medium and high conditions. Lastly, we aimed to extend the findings on Andrews et al. (2015) by collecting data on completion time. Particularly, we aimed to examine whether time could be used as a quantitative measure of familiarity, with familiar faces being sorted quicker than unfamiliar faces. In addition, if completion time could measure familiarity, perhaps the time data would also reveal an incremental benefit of ambient image exposure.

Method

Participants

The participants were 524 students from the University of Arkansas General Psychology participant pool. The participants were majority Caucasian (76.2%) and majority female (67.9%). The mean age of our sample was 19.26 years (SD = 2.08 years). A power analysis was conducted to determine the sample size. A small effect size was assumed to aim for high powered interaction effects and effects within the time data, which have not been previously measured in past studies. The power analysis found that we would need 84 participants per cell (6 cells total), for a total of 504 participants. Twenty participants were excluded and then replaced in the data set due to failing the manipulation check (n = 13), being unable to complete the task in time (n = 3), having prior familiarity with individuals pictured (n = 1), or due to researcher error (n = 3). An a priori decision was made to exclude and replace all participants who had a high misallocation error rate. Past studies have shown that misallocation errors are typically low (Andrews et al., 2015; Balas & Saville, 2017; Laurence et al., 2016; Redfern & Benton, 2017b). So, if participants made many misallocation errors, it is likely that they did not pay attention when sorting the faces. Specifically, if a participant had a misallocation error rate three standard deviations above the mean, they failed the manipulation check and were excluded.

Materials

The research team collected 65 images each from eight volunteers. All the volunteers were Caucasian females between the ages of 21-25. Each volunteer gave the research team permission to collect images from their social media page. The research team utilized the following guidelines when selecting photos: the target was no more than five years younger than their current age in each photo, the target was at least 18 years of age, and the target's face was clearly visible in each image. For the sorting task, the research team matched each volunteer for hair color, so that hair color could not be used as a cue on the following sort task. Four unique sort decks were created for this task. The research team ensured that all four decks were evenly distributed across trials. In addition, all four decks were evenly provided for the pretest and posttest sorting task.

All photos were cropped to reduce the amount of background information and to allow the face to be the focus of the image. The height of each photo was set to 3" and the width dimension was free to vary in order to not distort the image. All edits to the images were completed using either the app "paint" for a PC or "preview" for a Mac. Images were adjusted to 3" by adjusting the vertical pixels to 300 within each app. All images were colored photos. Only 20 of the 65 images collected from each volunteer were printed out and used during the sorting tasks. We used a random number generator to select the images that would be used for the learning tasks and the images that would be used for the sorting task. Each image was given an arbitrary number and then entered into the random number generator. The first 20 images that were selected by the generator were used as the images for the sorting task. This process was repeated for each set of 65 images of each identity. All final images were printed out and given a new arbitrary number from 1-20.

Procedures

Participants were brought into the lab one at a time to complete this study. The study is a mixed factor design consisting of three tasks; two face sorting tasks and a face learning phase.

For the first sorting task, all participants sorted images of unfamiliar identities. Then they viewed a face slideshow showing either five, 15 or 45 unique image exemplars of a target identity's face. For the second sorting task, participants were randomly chosen to sort unfamiliar identities or recently familiarized identities from the learning task. For both sorting tasks, the paradigm closely followed what was done in by Jenkins et al. (2011). After completing these tasks, each participant completed a demographics survey and were asked whether they previously knew any of the people shown during this study. This study took no longer than 30 minutes to complete.

Participants started with an unfamiliar face sorting task. This task featured photos of individuals that the participant did not know. For this task, participants were given 40 photos of faces and were told to sort the photos by identity. The pile contained 20 photos of two different people. Participants were not told how many identities were present in the pile. Participants were encouraged to perform the task with accuracy and speed. Additionally, participants were told that their performance would be timed. A previous study by Andrews et al. (2015) found that on average participants took about seven and a half minutes to complete an unfamiliar face sorting task, so participants in this study were given up to 10 minutes to complete this sorting task. After 10 minutes, participants were asked to make their final decisions. If they could not complete the task within this time period, their data were considered incomplete and not included in the final analysis. Only three participants were unable to complete the sorting task within 10 minutes.

After completing the unfamiliar face sorting task participants began the face familiarization phase. They were shown five, 15, or 45 images of four different identities. The four identities studied were matched for hair color and their images were intermixed together. Specifically, there were nine blocks in total for participants to view. For each block, participants viewed five images of each identity. The five images were presented sequentially and grouped by identity, so that within a block all images of a single identity were shown before images of the next identity were presented. Once five images of all four identities were presented, the block ended. The next block of images immediately appeared and cycled through all four identities. This repeated until all 9 blocks were shown. All images were shown sequentially and presented for 2 seconds each. Each identity was assigned a common six-letter name (Lizzie, Esther, Nicole, Ashley, Teresa, Summer, Brooke, Jasmin) to help participants distinguish between them. Participants viewed each identity for a total of 1 minute and 30 seconds. To match exposure time in each condition, each image was repeatedly shown a certain number of times so that participants would view each identity for a total of 1 minute and 30 seconds (i.e. if in the five-image condition, they would view these images nine times; if in the 15-image condition, they would view three unique blocks before images repeated). All images were displayed on a laptop computer and programed to show for the appropriate amount of time. Images were randomly selected to be shown in a random order.

Participants then completed another timed face sorting task. This sorting task either consisted of unfamiliar identities or identities from the face learning task. Similar to the first sorting task, this task consisted of 20 images of two different identities. Participants were unaware of how many identities were present and were simply asked to sort the images by identity. Speed and accuracy were encouraged on this task and all participants were made aware that they had up to 10 minutes to complete the task. This task was timed by the researcher. In addition, participants were asked to verify whether they recognize any of the sorted faces from the previous face learning phase. If they indicated "yes," they were asked to provide the name of the person featured in the pile. Participants were excluded and replaced if they indicated that they have prior familiarity with one of the identities featured in this study.

Results

We conducted separate analyses on the pile, time, and misallocation error data. Two participants were outliers in the familiar sorting task data for creating excessive piles (3 standard deviations greater than the mean of all familiar sorting tasks) and were therefore excluded from the final analysis. For all analyses, alpha was set at .05; however, all posttests had a corrected alpha using the Bonferroni adjustment. The pile generation data and misallocation data were positively skewed. In order to perform parametric tests, we performed a log10 transformation on the pile data. Some of the misallocation errors also fell at zero, so we added 1 to each data point and then performed a log10 transformation. For clarity, all means will be discussed using the original untransformed data.

Piles Generation Data

A 2 x 2 (Test [pretest, posttest] x Task [familiar sort, unfamiliar sort]) Analysis of Variance (ANOVA) was conducted to determine whether the first sorting task was significantly different than the second sorting task. Again, half of the participants were given a familiar sorting task for their second task (U-F), while the other half completed an unfamiliar sorting task (U-U). The ANOVA yielded a significant interaction effect, $F(1, 499) = 32.22$, $p <.001$, $\eta_p^2 = .06$. Pretest scores were not significantly different for the U-F ($M = 6.97$, $SD = 4.65$) group and the U-U ($M = 6.91$, $SD = 4.43$), $p = .89$. However, posttest scores differed, indicating that participants in the U-F group ($M = 4.36$, $SD = 2.99$) created fewer piles than participants in the U-U ($M = 5.85$, $SD = 3.92$) group, $p <.001$. Both groups showed improved scores from pretest to posttest, (U-U, $p < .001$ & U-F, $p < .001$).

Next a 2 x 3 (Task [familiar sort, unfamiliar sort] x Images Studied [5, 15, 45]) ANCOVA was conducted on the pile generation data. The results are displayed in *Figure 1*. Pile sorting performance on the pretest task was used as a covariate and found to significantly account for performance on the posttest task, $F(1,498) = 259.78$, $p < .001$, $\eta_p^2 = .343$. The results showed a main effect of task, in which participants created significantly fewer piles when sorting familiar faces ($M = 4.36$, $SD = 2.99$) than when sorting unfamiliar faces ($M = 5.85$, $SD = 3.92$), $F(1, 498) = 46.22$, $p < .001$, $\eta_p^2 = .09$. A main effect of images studied was also found, $F(2,498) = 4.08$, $p = .02$, $\eta_p^2 = .02$. The five-image condition ($M = 5.44$, $SD = 3.39$) was significantly different from the 45-image condition ($M = 4.74$, $SD = 3.13$), p = .01. There was no interaction effect, $F(2, 498) = 1.88$, $p = .15$. However, to parse apart the critical question of whether exemplar exposure lead to better performance on the familiar sorting task, a follow up one-way ANOVA was conducted on the familiar data. The results revealed that the number of images shown did have a significant effect on the piles generated, $F(2, 248) = 5.08$, $p = .003$, $\eta_p^2 = .04$. Follow up posttests revealed that participants who studied five unique images ($M = 4.95$, $SD = 2.90$) created significantly more piles than those who studied 15 images ($M = 4.18$, $SD = 3.25$, $p = .03$) or 45 images ($M = 3.95$, $SD = 2.73$, $p = .003$). However, there was no significant difference between the 15-image condition and the 45-image condition, $p > .99$. A one-way ANOVA of the unfamiliar data did not show an effect of number of images studied, $F(2, 249) = 1.1$, $p = .34$.

Next, separate single-sample t-tests were conducted to determine whether the group means of the familiar data were significantly different from 2. If the group means were no different from 2, it would suggest that participants accurately recognized the two identities present in the sort task. The results showed that the five-image condition, $t(83) = 12.25$, $p < .001$, the 15-image condition, $t(82) = 9.45$, $p < .001$, and the 45-image condition, $t(82) = 10.34$, $p < .001$, all resulted in piles significantly greater than two. In short, participants did not achieve perfect performance on any of the familiar face sorting conditions.

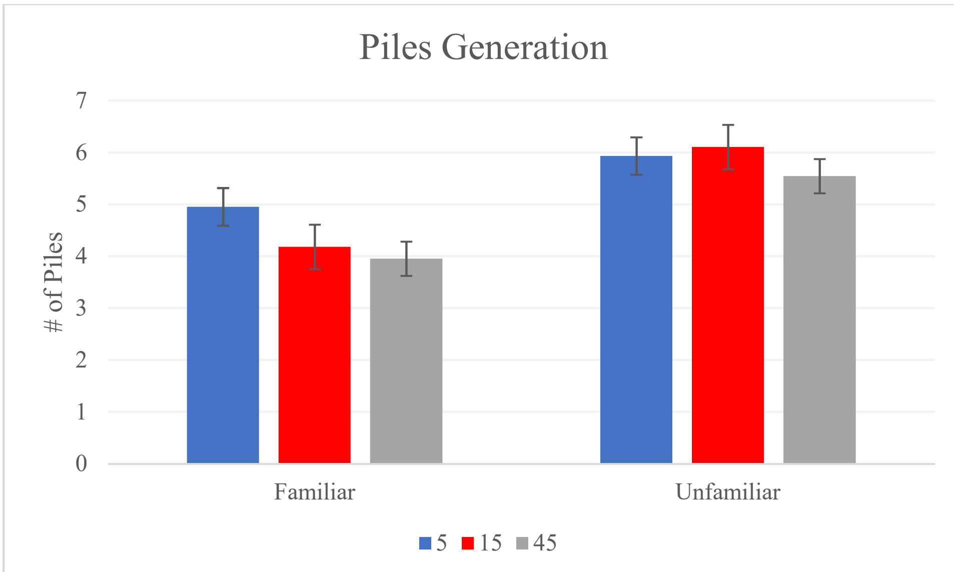

Figure 1. A bar graph of the mean number of piles generated by each group. Groups are divided by the number of images they studied and whether they received a familiar or unfamiliar face sorting task. The error bars capture standard error for each group.

Another group of single-sample t-tests was conducted to determine whether the group means of the familiar data were significantly different from 4. If the group means were no different from 4, it would suggest that the participants may have been led to create four piles after learning four new identities. The five-image condition, was not significantly different from 4, $t(83) = 1.00$, $p = .320$. However, both the 15-image condition, $t(82) = -2.72$, $p = .008$, and the 45-image condition, $t(82) = -3.39$, $p = .001$, were significantly different from four. One major concern of this project was whether performance might have been impacted by expectation (the expectation that four people will be present in the sorting task if four people were present in the learning phase). *Figure 2* displays the distribution of piles created among participants and can help inform whether participants were solely impacted by expectations. Looking at the frequency of piles generated, most participants created two, three, or four piles.

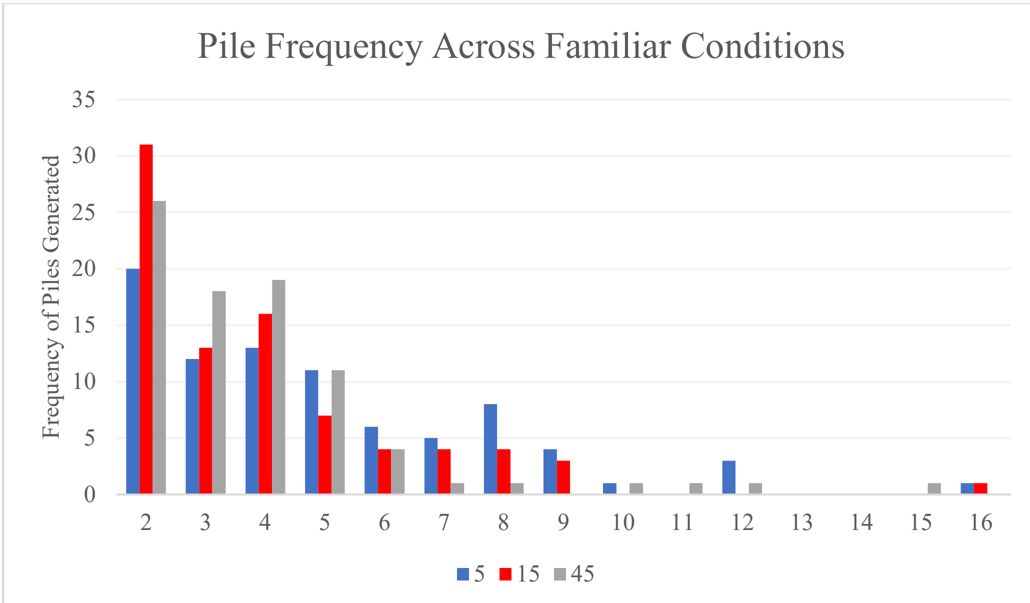

Figure 2. A frequency distribution of the number of piles generated by participants in the familiar face sorting task condition.

As mentioned earlier, four unique decks were created for this task. Each participant sorted up to two of these decks; one at pretest and one at posttest. Each deck featured 20 photos of two unique identities and all decks were evenly counterbalanced across conditions. Each deck was arbitrarily given a label of one through four. These decks may have differed in terms of difficulty to sort, so an additional analysis was conducted on the pile data to determine whether the decks used during the task interacted with any of the independent variables. To ensure that the four sort decks did not affect the results, a 4 x 2 x 3 (Sort Decks [first, second, third, fourth] x Familiarity [familiar sort, unfamiliar sort] x Images Studied [5, 15, 45]) ANOVA was conducted. The results revealed that there was a main effect of sort deck, $F(3, 480) = 10.97$, $p < .001$, $\eta_p^2 = .06$. The third deck ($M = 5.60$, $SD = 3.32$) was sorted into more piles on average than the first (M

= 5.37, *SD* = 4.01, *p* = .01), second (*M* = 4.52, *SD* = 3.20, *p* < .001), and fourth deck (*M* = 4.95, *SD* = 3.60, *p* < .001). No other decks were significantly different from one another. In addition, no interaction effects were observed between the sort decks and other factors.

Time Data

Time differences between familiar and unfamiliar sorting tasks have not yet been recorded by past studies. We hypothesized that participants would be able to more quickly sort familiar faces than unfamiliar faces. Finding support for this hypothesis would suggest that time data can be used as another measure of familiarity on face sorting paradigms. All time data are reported in seconds. First a 2 x 2 (Test [pretest, posttest] x Task [familiar sort, unfamiliar sort]) ANOVA was conducted to observe whether time data could reflect familiarity with the test stimuli. There was a significant interaction effect, $F(1, 503) = 50.98$, $p < .001$, $\eta_p^2 = .09$. Pretest completion times were not significantly different for participants in the U-U (*M* = 331.69 sec, *SD* = 159.38 sec) and U-F (*M* = 319.94 sec, *SD* = 152.86 sec) condition, *p* = .40. However, posttest completion times were significantly faster for the U-F condition (*M* = 227.18 sec, *SD* = 134.40 sec) than the U-U condition (*M* = 311.64 sec, *SD* = 142.52 sec), *p* < .001. In addition, participants in both the U-U condition and the U-F condition completed the sorting task significantly faster at posttest than pretest (*p* = .006 & *p* < .001, respectively).

Next a 2 x 3 (Task [familiar sort, unfamiliar sort] x Images Studied [5, 15, 45]) ANCOVA was conducted on the time data. The results are displayed in *Figure 3*. Pretest completion time was used as a covariate and was found to significantly account for posttest completion time, $F(1, 498) = 512.18$, $p < .001$, $\eta_p^2 = .51$. There was a main effect of task, in which familiar items (*M* = 230.89 sec, *SD* = 134.40 sec) were sorted significantly faster than unfamiliar items (*M* = 307.96 sec, *SD* = 142.52 sec), $F(1, 498) = 78.97$, $p < .001$, $\eta_p^2 = .14$. There was also a main effect of images studied, $F(2, 498) = 3.470$, $p = .03$, $\eta_p^2 = .01$. Only the five-image ($M = 284.83$ sec, $SD = 136.12$ sec) condition was sorted significantly slower than the 45-image ($M = 257.87$ sec, $SD = 154.85$ sec) condition, $p = .03$.

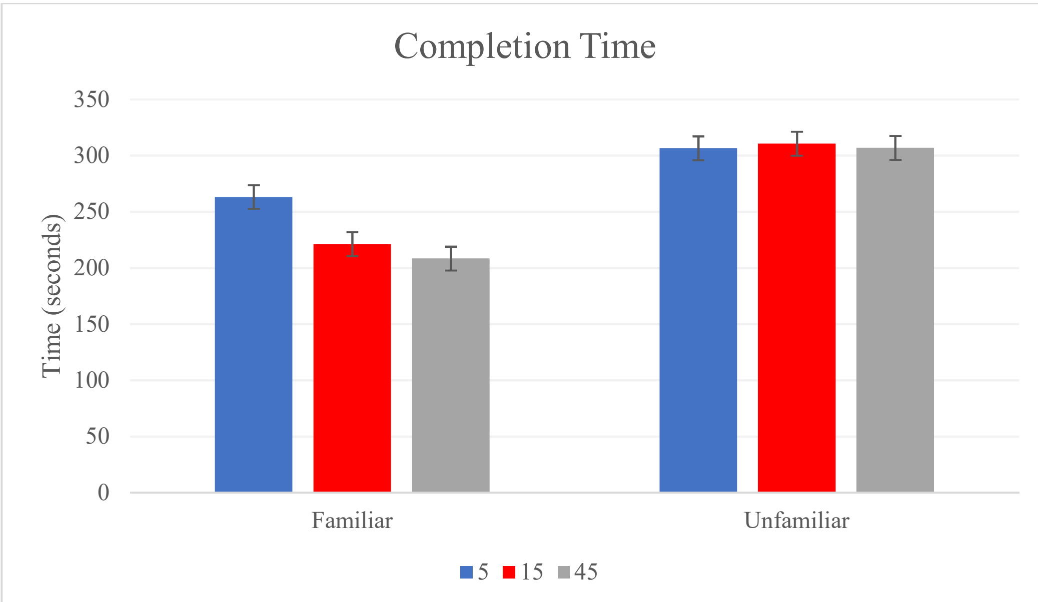

Figure 3. A bar graph of the mean completion time of each group. Groups are divided by the number of images participants studied and whether they received a familiar or unfamiliar face sorting task Time is reported in seconds. The error bars capture standard error for each group.

The interaction was significant as well, $F(2, 498) = 3.89$, $p = .02$, $\eta_p^2 = .02$. Among the unfamiliar data, no group was significantly different from one another regardless of how many images they studied. However, for the familiar data, the five-image group ($M = 263.17$ sec, $SD = 120.77$ sec) sorted piles significantly slower than both the 15-image group ($M = 221.19$ sec, $SD = 128.03$ sec), $p = .02$, and the 45-image group ($M = 208.32$ sec, $SD = 150.95$ sec), $p = .001$. The 15-image group did not significantly differ from the 45-image group, $p > .99$. In addition, when matched for the number of images studied, all familiar conditions were completed faster than the unfamiliar conditions (five-image condition ($M = 5.93$, $SD = 3.77$), $p = .004$; 15-image condition ($M = 6.10$, $SD = 4.60$), $p < .001$; 45-image condition ($M = 5.54$, $SD = 3.32$), $p < .001$).

Misallocation Errors

Misallocation errors are measured according to how frequently a participant places images of a different people within the same pile. Similarly, to past studies, the identity that is present the least within a pile was counted as a misallocation error (Andrews et al., 2015; Balas & Saville, 2017; Jenkins et al., 2011; Laurence et al., 2016; Redfern & Benton, 2017b). For example, if a pile contains four photos of "Identity A" and two photos of "Identity B," we would count this as two misallocation errors. On average participants made very few misallocation errors. In fact, consistent with past studies, during this study, participants made an average of 1.77 misallocation errors when collapsed across sorting tasks. To further explore misallocation errors by condition we conducted a 2 x 3 (Task [familiar sort, unfamiliar sort] x Images Studied [5, 15, 45]) ANCOVA. Pretest misallocation error rates were used as the covariate and were found to significantly account for posttest misallocation errors, $F(1, 498) = 46.42$, $p < .001$, $\eta_p^2 = .09$. There was also a main effect of familiarity, in which participants sorting familiar faces ($M = 1.11$, $SD = 1.87$) made significantly fewer misallocation errors than participants sorting unfamiliar faces ($M = 2.43$, $SD = 3.68$), $F(1, 498) = 30.99$, $p < .001$, $\eta_p^2 = .06$. The main effect of number of images studied, $F(2, 498) = .03$, $p = .97$, and the interaction effect, $F(2, 498) = .40$, $p = .67$, was not significant. To keep consistent with the previous analyses, a follow up one-way ANOVA was conducted to observe whether the number of images studied had any effect on the familiar face sorting condition. The results showed that the number of images studied had no significant effect on misallocation error rate among participants completing the familiar face sorting task, $F(2\ 248) = .28$, $p = .76$. These results showed that misallocation errors are not as sensitive to our manipulation as time and pile data seemed to be.

Discussion

Of the three hypotheses generated, we found support for one and partial support for another. First, familiarity was found to predict time data. Particularly, participants could more quickly sort face images that were familiar than face images that were unfamiliar. This result has the potential of helping to interpret future studies that use this face sorting paradigm. Next, we found some evidence that as participants studied more images, they created fewer piles and completed the sorting task faster. This was only true for the five-image condition when compared to both higher-numbered conditions. No improvement was found in neither the time data nor pile data when comparing the 15-image condition to the 45-image condition. This finding is important because past studies have often only compared a large image exemplar group to a smaller image exemplar group. However, by adding a critical intermediate group, it becomes clear that the incremental benefit of studying more ambient images quickly becomes less pronounced after participants view a large set of images. The results showed that an additional 10 images boosted performance significantly (fewer piles and faster completion times) if participants were shown five versus 15 images. However, this same improvement was not observed between 15 and 45 images, though an additional 30 images were shown. The incremental benefit of ambient images becomes less prominent as exemplar variance increases. Lastly, our manipulation was unable to produce perfect performance. Even when shown 45 ambient images of a person's face, not all participants could perform the task perfectly. This result suggests that more than just visual familiarity is needed to mimic real-life familiarity in the lab.

This is the first study to find evidence that time data can be used as a quantitative measure of familiarity on Jenkins et al.'s (2011) face sorting paradigm. Particularly, we manipulated familiarity by providing a face learning phase. The second face sorting task either included identities from the face learning phase or completely novel identities. If asked to sort faces that they recently studied, participant should be more familiar with that identity than an identity that they never seen before. The time data results reveal that the familiarity manipulation did indeed work. Participants completed the sorting task faster when working with recently familiarized faces than when working with unfamiliar faces. Past studies have used somewhat different methods to analyze the results of this face sorting paradigm (Andrews et al., 2015; Balas & Saville, 2017; Jenkins et al., 2011; Laurence et al., 2016; Redfern & Benton, 2017b). Therefore, using time data might serve as a unified method of capturing familiarity effects to allow for easier comparisons between studies. Also, it may help provide further understanding of why performance differs on this task. A future study might want to assess whether time data can reveal differences in performance when sorting recently familiarized faces versus familiar celebrity faces to further validate the usefulness of time data on this task.

Another major finding is that there appears to be a minuscule improvement on this task after viewing 15 ambient images of a novel identity. In this study, showing an additional 30 novel images of an identity did not result in a significant improvement in performance. One implication of this result is that there is a limit to visual familiarity. After showing a certain number of images of a person's face, participants may be unable to become any more familiar with a face through exposure to image exemplars alone. This idea is further supported by this study because participants also did not show an ability to reach perfect performance, as the result of being shown 45 exemplar images. Perhaps 45 images are too low a number to show any additional statistical benefit. However, even Murphey et al. (2016) failed to achieve perfect recognition performance after passively having participants view 96 images of eight identities.

Also, in the current study, the number of exemplars studied was twice the size of the images the participants had to sort at test. Thus, it is expected that participant would have seen enough exemplar variance to account for the variability of the images seen during the familiar sorting task. Being unable to achieve perfect performance on this task might suggest that there is some limit to the benefit of ambient image exposure. In Jenkins et al. (2011), participants completed the task perfectly when sorting familiar celebrity faces, suggesting that perfect performance on this sorting task reflects strong familiarity with the target face. If this did not occur from our manipulation, then perhaps our participants are not yet fully familiar with the studied identities. Particularly, people likely need more than just image exposure to become familiar with a face.

Past studies have suggested that there may be some limit to ambient images. As mentioned in the introduction, Andrews et al., 2015 theorized a limit to the benefit of ambient images after finding no difference in recognition rates between participants exposed to face images in a free-sort condition versus those in an informed-sort condition. This study used the same face sorting paradigm but used this paradigm as a study phase. The participants who were informed that there was two faces present had the opportunity to integrate all 20 images as representing one identity. While the participants in the free sort condition on average had fewer images integrated as representing one identity. Despite this, the authors found no difference between groups on a follow-up face matching task. The authors conclude that perhaps both groups created a sufficient representation of each identity despite the differences in the number of successful piles they created of each identity in the study phase (Andrews et al., 2015). In addition, as mentioned earlier, another study used computer software to create face averages by combining ambient images of a single identity's face (Jenkins & Burton, 2011). They found that after combining 12 face images, the average image become relatively stable and would not change much upon adding additional images. The authors conclude that human facial recognition may behave in a similar fashion, if humans also construct a prototypical refence of an identity when learning several face images (Jenkins & Burton, 2011).

Collectively these findings suggest that it is necessary to extend past exemplar variance in order to understand face familiarity. Our findings as well as past findings can be better understood through the lens of a theoretical model constructed by Bruce and Young (1986). In their model, they suggested that when recognizing a face, several codes become activated—pictorial, structural, name, semantic, and expression. Often when discussing ambient images, pictorial and structural codes are heavily considered. Pictorial codes refer to the features of the actual image of a face (lighting, angle, shadowing, etc.), while structural codes refer to the features of a face that remain consistent across image exemplars. However, Bruce and Young's model highlighted that semantic codes, name codes, and expression codes are also activated when viewing a face. By only training participants on structural codes (pictorial codes are less important in this task since we did not test participants on the exact images they studied), we neglected to allow participants to develop additional codes that may aid in their recognition ability. For example, identity-specific semantic codes include information such as context, episodic memories, and general knowledge of that person (Bruce & Young, 1986). In this study, participants could not benefit from this code when sorting faces. In fact, these identities were learned in the same context and likely had some overlap in the episodic experiences associated with them. We can expect that no unique identity-specific semantic code was associated with any of the faces learned in our task. As a result, this might have impacted our participants' ability to identify all images of our target identity during the sorting task. Perhaps finding a method to encourage the development of identity-specific semantic codes could lead to achieving perfect performance on this task.

Past neuroscience studies have also found evidence to support the importance of accessing additional codes when recognizing faces. A literature review by Natu & O'Toole (2011) found that having stronger neural representations of a face indicates greater familiarity. This literature review highlighted neuroscience articles that demonstrated the activation of several regions of cortex associated with emotions and memories when viewing familiar faces. Natu and O'Toole also made the distinction that faces can be visually, personally, or impersonally (famous persons) familiar. A recent study found evidence of increased neural activity when viewing personally familiar faces, that is not observed when viewing celebrity faces or unfamiliar faces (Wiese et al., 2019). This study found EEG evidence that support a differentiation between personally familiar faces, as well as less personally familiar faces (i.e., professors) and unfamiliar faces. However, this effect is not seen when comparing celebrity faces to unfamiliar faces. Specifically, the authors found a "sustained-familiarity effect," characterized as a prolonged N250 ERP. This sustain-familiarity effect was only found when comparing highly familiar face and less personally familiar faces to unfamiliar faces. This finding suggests that our ability to recognize faces is highly variant and familiarity falls on a spectrum. There are different levels of familiarity, which may be partially dependent on the many different codes that one associates to any individual identity.

Perhaps the real-life experience that the participants have with personal familiar faces, regardless of closeness, enhances familiarity effects. It would be interesting to see if this effect could be captured with behavioral data. There are different levels of familiarity. As a face becomes more familiar, it is likely that there are more codes that can be accessed to help make an identity judgment. A next step of facial recognition literature is to integrate our knowledge of ambient images into better understanding the more complex interactions involved with naturally becoming familiar with a new person. Particularly, increasing behavioral evidence of what has been suggested through neuroscience evidence could enhance what we know about face recognition and the process of how a face becomes familiar.

Alternatively, the results of this study could be explained by the participant's expectancy. Though not all participants were led to believe that there were four people present in the sorting task as a result of viewing four identities in the learning phase, there is evidence that several participants may have been affected by this manipulation. In follow up studies, it will be important to find methods to prevent participants from using alternative strategies when completing this sorting task. Perhaps adding random unfamiliar faces to the sort deck can help break any expectation about the number of piles the participant should sort. Despite this potential downfall, these data still yield meaningful results. In short, the results reveal that even with 45 image exemplar exposures, participants still did not have enough mastery of the studied identity's face to successfully complete the task. If the participant had mastery of the familiarized identity's face, we would expect to see that they were able to override any expectancy about how many piles they believe they should generate. For example, if this paradigm included celebrity or personally familiar faces, it is likely that participants would be able to complete this task perfectly, regardless of any expectation of how many piles they should generate.

The current study contributes to face recognition literature in two major ways. First providing evidence of the usefulness of time data can help contribute to interpreting the results of future studies examining facial recognition. In addition, this study and the subsequent discussion highlights some potential downfalls of focusing solely on exemplar variance as we try to understand face familiarity. Future studies might attempt to determine whether there are other methods of recreating familiarity in the lab, perhaps by increasing the relevance of other recognition cues. Typically, missing persons advisories will feature several images of a missing person in hopes of increasing the chances of an identification. Past studies have even found evidence of the utility of presenting multiple images in missing person or wanted person notices (Lampinen, Curry, & Erikson, 2016; Sweeney & Lampinen, 2012). However, if visual familiarity only leads to partial improvements in face recognition performance, it may be worth determining other methods of artificially creating familiarity. By providing additional recognition codes, perhaps we can continue to improve the probability that a missing person can be identified by a stranger.

Conclusion

The current study provide evidence regarding whether an unfamiliar face can become familiar through ambient image exposure alone. Based on the results, more than just visually familiarity is likely needed in order to mimic real-life familiarity in the lab setting. Previous literature has repeatedly shown that exposing participants to multiple ambient images of a face can lead to improved recognition rates and improved familiarity. However, there was a gap in the literature of whether there might be a limit to the benefit of ambient images. The current study attempts to provide evidence for some incremental growth in recognition by having participants view either five, 15 or 45 images of an unfamiliar person prior to performing a sorting task. However, performance did not significantly improve when show 45 images versus only 15, suggesting a limit to visual familiarity. This study also found evidence that completion time on sorting tasks can serve as an additional quantitative measure of familiarity. This finding is worth replicating with other face stimuli, such as celebrity faces or personally familiar faces. In the original Jenkins et al. (2011) study, participants were able properly sort face images when given familiar face stimuli but not when given unfamiliar face stimuli. The implications suggest that overall people are poor at predicting the variability of face images of unfamiliar identities because they often consider highly variant images of an individual as two separate identities. Exemplar variance is necessary to better tolerated the variability of face images of a novel identity. However, once a person is better able to tolerate ambient image variability, they may be able to benefit from learning additional cues about a face as well. An interesting next step for face recognition literature is to begin to integrate knowledge on exemplar variance and other codes of facial recognition in order to increase our understanding of how a face becomes familiar.

References

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *The Quarterly Journal of Experimental Psychology, 68*(10), 2041-2050. doi:10.1080/17470218.2014.1003949

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research, 141*, 228. doi: 10.1016/j.visres.2016.12.005

Besson, G., Barragan-Jason, G., Thorpe, S. J., Fabre-Thorpe, M., Puma, S., Ceccaldi, M., & Barbeau, E. J. (2017). From face processing to face recognition: Comparing three different processing levels. *Cognition, 158*, 33-43. doi:10.1016/j.cognition.2016.10.004

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PloS One, 11(2)*, e0148148. doi:10.1371/journal.pone.0148148

Bruce, V. (1994). Stability from variation: The case of face recognition the MD Vernon memorial lecture. *The Quarterly Journal of Experimental Psychology*, *47*(1), 5-28.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5(4)*, 339–360. doi: 10.1037/1076-898X.5.4.339

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7(3)*, 207–218. doi: 10.1037/1076-898X.7.3.207

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77(3)*, 305–327. https://0-doi-org.library.uark.edu/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science, 40(1)*, 202-223. doi:10.1111/cogs.12231

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195(4275)*, 312-314. doi:10.1126/science.831281

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology, 62(9)*, 1716. doi:10.1080/17470210902811249

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115(2)*, 107-117. doi:10.1037/0096-3445.115.2.107

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *The Quarterly Journal of Experimental Psychology, 69(1)*, 1-10. doi:10.1080/17470218.2015.1017513

Ellis, H.D., Shepherd, J.W. & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception, 8*, 431 –439.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121(3)*, 313–323. DOI: 10.1016/j.cognition.2011.08.001

Kramer, R. S., & Ritchie, K. L. (2016). Disguising superman: How glasses affect unfamiliar face matching. *Applied Cognitive Psychology*, *30(6)*, 841-845.

Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision, 15(4)*, 1-9. doi:10.1167/15.4.1

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology, 11(3)*, 211-222. doi:10.1002/(SICI)1099-0720(199706)11:3<211::AID-ACP430>3.0.CO;2-O

Lampinen, J. M., Curry, C. R., & Erickson, W. B. (2016). Prospective person memory: The role of self-efficacy, personal interaction, and multiple images in recognition of wanted persons. *Journal of Police and Criminal Psychology*, *31(1)*, 59–70. https://0-doi-org.library.uark.edu/10.1007/s11896-015-9164-7

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. *British Journal of Psychology, 107(2)*, 374-388. doi: 10.1111/bjop.12147

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34(4)*, 865-876. doi: 10.3758/BF03193433

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied, 14(4)*, 364-372. doi: 10.1037/a0013464

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology, 27(6)*, 700-706. doi:10.1002/acp.2965

Menon, N., White, D., & Kemp, R. I. (2015). Variation in photos of the same face drives improvements in identity verification. *Perception, 44(11)*, 1332-1341. doi: 10.1177/0301006615599902

Mileva, M., & Burton, A. M. (2018). Smiles in face matching: Idiosyncratic information revealed through a smile improves unfamiliar face matching performance. *British Journal of Psychology, 109(4)*, 799-811. doi: 10.1111/bjop.12318

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology. Human Perception and Performance, 41(3)*, 577-581. doi: 10.1037/xhp0000049

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology*, *102(4)*, 726–747. https://0-doi-org.library.uark.edu/10.1111/j.2044-8295.2011.02053.x

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition, 128(1)*, 56-63. doi: 10.1016/j.cognition.2013.03.006

Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications, 3(1)*, 1-9. doi: 10.1186/s41235-018-0110-y

Redfern, A. S., & Benton, C. P. (2017a). Expression dependence in the perception of facial identity. *I-Perception, 8(3)*, doi: 10.1177/2041669517710663 2041669517710663.

Redfern, A. S., & Benton, C. P. (2017b). Expressive faces confuse identity. *I-Perception, 8(5)*, 2041669517731115. doi: 10.1177/2041669517731115

Redfern, A. S., & Benton, C. P. (2018). Representation of facial identity includes expression variability. *Vision Research*. doi: 10.1016/j.visres.2018.05.004

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *The Quarterly Journal of Experimental Psychology*, *70(5)*, 897–905. doi: 10.1080/17470218.2015.1136656

Ritchie, Kay L., Robin S. S. Kramer, and A. M. Burton. (2018). What makes a face photo a "good likeness"? *Cognition*, *170*, 1-8. doi: 10.1016/j.cognition.2017.09.001

Sweeney, L. N., & Lampinen, J. M. (2012). The effect of presenting multiple images on prospective and retrospective person memory for missing children. *Journal of Applied Research in Memory and Cognition*, *1(4)*, 235–241. https://0-doi-org.library.uark.edu/10.1016/j.jarmac.2012.08.001

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS One, 9(8)*, e103510. doi: 10.1371/journal.pone.0103510

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B*, *282*(1814), doi: 10.1098/rspb.2015.1292.

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y. X., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. *Psychological Science*, *30(2)*, 261–272. https://0-doi-org.library.uark.edu/10.1177/0956797618813572

Valentine, T. (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition, The Quarterly Journal of Experimental Psychology Section A, *43(2)*, 161-204. doi: 10.1080/14640749108400966

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81(1)*, 141-145. doi:10.1037/h0027474

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M, & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14(6)*, 737-746.

Appendix

UNIVERSITY OF ARKANSAS

To:	Nia I Gipson
From:	Douglas James Adams, Chair IRB Committee
Date:	09/04/2018
Action:	**Exemption Granted**
Action Date:	09/04/2018
Protocol #:	1807134409
Study Title:	Make Unfamiliar Faces Familiar

The above-referenced protocol has been determined to be exempt.

If you wish to make any modifications in the approved protocol that may affect the level of risk to your participants, you must seek approval prior to implementing those changes. All modifications must provide sufficient detail to assess the impact of the change.

If you have any questions or need any assistance from the IRB, please contact the IRB Coordinator at 109 MLKG Building, 5-2208, or irb@uark.edu.

cc: James Michael Lampinen, Investigator