[1]   The Effect of Image Presentation Rate on Person Identification

[2]   Carlos M. Ibaviosa[1] & Rachel A. Searston[1,2]

[3]   [1] University of Adelaide

[4]   [2] University of Adelaide

[5]   Author Note

[6]   Add complete departmental affiliations for each author here. Each new line herein

[7] must be indented, like this line.

[8]   Enter author note here.

[9]   Correspondence concerning this article should be addressed to Carlos M. Ibaviosa, .

[10] E-mail: carlos.ibaviosa@gmail.com

Abstract

Our ability to recognise complex images across contexts depends on our exposure to similar instances. For example, despite much natural variation, it is easier to recognise a new instance of a familiar face than an unfamiliar face. As we encounter similar images, we automatically notice structural commonalities and form a representation of how the image generally looks, even when each image is presented rapidly (i.e., several milliseconds each). However, it is not clear whether this process allows us to better identify new instances of an image compared to assessing single images for a longer duration. Across two experiments, I tested observers' person recognition ability when presented with rapid image streams at varying rates compared to a single image. Experiment 1 compares performance between upright and inverted faces. Experiment 2 compares performance between fingerprints from the same finger and from the same person more generally. My results suggest that viewing images rapidly is better than single images when identifying faces, but not fingerprints; and that people better recognise upright compared to inverted faces, but are similar in both fingerprint conditions. I discuss the theoretical implications of these results, as well as some practical implications in security and forensic contexts.

Keywords: Visual cognition, recognition, gist perception, ensemble coding, face processing, fingerprint analysis

Word count: X

The Effect of Image Presentation Rate on Person Identification

Contents

```
# Seed for random number generation
set.seed(42)
knitr::opts_chunk$set(cache.extra = knitr::rand_seed)
```

Significance Statement

Forensic examiners in various fields are regularly required to make identification decisions based on complex, unfamiliar images – such as a stranger's face, or a stranger's fingerprint – often based on a single comparison photo, or a limited number of comparison photos. While much evidence suggests that recognising a new image would benefit from viewing multiple different examples of that image beforehand, fewer studies have explored whether it is more beneficial to view several comparison photos quickly, or a single comparison photo for a longer duration, if given a limited time to make the identification. If quickly processing several images leads to greater image recognition, then a similar approach could be used to better allocate time resources, or streamline training in many forensic identification disciplines. In this study, we tested this idea under various different conditions, using face (Experiment 1) and fingerprint (Experiment 2) stimuli, with novice participants. While we speculated on many possible constraints when applying this methodology under different conditions, we generally found that while there was an advantage to quickly viewing several images, this advantage was more pronounced with more familiar image categories, and was slightly affected by image specificity.

Introduction

[Fix up expression]

[Make sure all links to OSF pages are working]

[Make sure all references are done...]

Our ability to correctly categorise an object or image seems to depend on how much experience we have in viewing similar kinds of objects in the first place. For example, the

prototype theory of categorisation suggests that when categorising an object, we compare it to the typical representation of similar objects in our long-term memory and categorise it accordingly (reference) . Similarly, the exemplar theory of categorisation suggests that, when recognising an object, we compare it to our memories of specific objects within a particular category that we have accumulated in the past (references) , and search for similarities. Due to this reliance on similar prior experiences, it tends to be more difficult to categorise objects that we do not see very often, because we are not familiar with how these objects may vary under different contexts, or are unaware of the more stable, average characteristics among these objects that may facilitate categorisation (reference) . On the particular level, for example, a substantial body of literature has focused on the role of familiarity in individual face recognition. Indeed, trying to identify a stranger's face proves much more difficult than identifying a friend's face or a celebrity's face, because we do not know what a stranger's face typically looks like and how it varies across contexts, and may mistake simple variations in lighting or hairstyle for complete changes in identity (references) . This is not the case for familiar faces, where we can remember their stable facial features across contexts, and can easily recognise those features even in a new environment (references). However, even if we do not have exposure to various instances of the same object, evidence suggests that our cumulative experience in viewing various instances of the broader category can still yield an advantage. Fingerprint experts, for example, can better identify two unfamiliar fingerprints compared to novices because their vast experience with fingerprints generally allows them to better understand how fingerprints vary.

If our ability to effectively recognise and categorise different objects, both on an individual and categorical level, is assisted by our understanding of the commonalities between members of a particular category, how then do we make sense of these commonalities? One related explanation is "ensemble coding", which allows us to glean the average properties of a range of similar stimuli and automatically make sense of the common characteristics in our environment (references) . However, while the previous studies in

117 identification and categorisation may suggest that learning regularities among a category

118 depends on having ample exposure to each individual instance - for example, face recognition

119 studies often give participants several seconds to learn new faces (references), and fingerprint

120 (reference) experts will have spent hours in cumulatively viewing objects in their domain of

121 expertise - research in ensemble coding suggests that committing each instance to visual

122 memory over time may not even be necessary. In fact, many studies using the rapid serial

123 visual presentation (RSVP) methodology, where a series of similar images are presented for

124 several milliseconds each one after the other, have shown that we can automatically compute

125 the average representation of all of the images - despite not being able to process any

126 individual image. This finding has been replicated for when participants focus on simple

127 stimuli (e.g., average circle size; reference), complex stimuli (e.g., average facial expression;

128 reference), and even when the RSVP stream is not the main focus of the experiment

129 (reference). However, while ensemble coding is very robust to task demands, and it seems

130 intuitively linked to how we become familiar with a set of images, no studies seem to have

131 established whether presenting unfamiliar images in an RSVP stream can help to identify

132 new images of the same category. The current study, therefore, asks whether rapidly viewing

133 the gist of several images can improve novices' ability to identify unfamiliar objects

134 compared to carefully assessing the details of a single image, using strangers' faces and

135 fingerprints as visual stimuli.

136   Previous research has suggested that exposing novices to several instances may better

137 simulate expertise than only assessing single images (double check if this is true - Thompson

138 & Tangen, 2014), and so this may be a powerful methodology to do so. However, previous

139 research has also suggested that visual expertise has its limits (Bukach, Phillips & Gauthier,

140 2010; Diamond & Carey, 1986; see and include Searston & Tangen, 2017), and so we will also

141 explore the possible constraints to this methodology when considering other variables that

142 may influence recognition in these contexts.

- talk about holistic processing??

- image variability

- working memory demands

The Current Study

The present research examines whether viewing an RSVP stream of images at varying rates can better facilitate object recognition compared to viewing a single image, when presented for an equal duration of time. While several studies on face recognition have already suggested that it is better to view more photos of a person compared to fewer photos (e.g., Murphy et al., 2015), no studies seem to have directly compared whether it is better to carefully assess the details of a single image, or the get the general gist of several images rapidly, when making an identification - and so our study will be the first to do so. Across two experiments, we presented participants with complex, unfamiliar images representing the same person (i.e., a stranger's face or fingerprint), as either single images, or as RSVP streams at varying rates (i.e., two, four, and eight images per second) for a total of eight seconds. In each trial they were asked whether they viewed images from the same or different category to the test image (e.g., "Is this the same person?"). Based on previous research, we expect that recognition performance will increase as participants view more images per second, given that this would allow them to create richer ensemble representations compared to other conditions. In essence, viewing more images per second may allow participants to become "more familiar" with the unfamiliar stimuli presented, making it easier to recognise any common features shared with the test image and make an appropriate identification or rejection.

In Experiment 1, participants view the RSVP streams *before* viewing the test image, as previous research suggests (references) that it is the accumulation of multiple previous exemplars that facilitates face recognition. We will also explore whether recognition in these different conditions is affected by familiarity with the general stimulus class, by presenting

169 the faces as upright and inverted images. Previous research in visual recognition has

170 suggested that we are much better at recognising upright faces compared to inverted faces

171 (see Rossion, 2008 for a review) - possibly due to our disproportionate experience in viewing

172 upright faces everyday (references), an innate ability to do so more efficiently (references), or

173 a combination of both. In manipulating face inversion in this experiment, we can examine

174 the influence of ensemble coding in recognition tasks not only on an individual level of

175 familiarity, but on a group level: if ensemble coding tends to be automatic and accurate

176 when viewing simple stimuli (e.g., circle size; references) and complex familiar stimuli (e.g.,

177 upright faces; references), would it operate the same way when viewing complex *unfamiliar*

178 stimuli (i.e., inverted faces)? It is possible that inverted faces may not share the same benefit

179 as upright faces, as the difficulties in processing inverted faces holistically (Tanaka &

180 Simonyi, 2016; see also Rossion, 2008 for a review) may make it more difficult to process the

181 gist of each image in the stream. However, given how automatic ensemble coding is in a

182 variety of tasks with a variety of stimuli, we believe that our methodology could nevertheless

183 exert a positive influence with face recognition (but see Haberman et al., 2009; Haberman,

184 Lee, & Whitney, 2015; and Leib, Puri, Fischer, Bentin, Whitney, & Robertson, 2012, in

185 relation to ensemble coding generally - what have these studies said???). In fact, we predict

186 that any benefit derived from ensemble coding may actually be more pronounced when

187 viewing inverted compared to upright faces, given that our existing advantage for upright

188 face-matching may limit how beneficial this methodology may be for upright faces relative to

189 inverted faces, which do not share the same constraints. Additionally, we expect that upright

190 face recognition may also be constrained by the flashed-face distortion effect (FFDE), where

191 faces in an RSVP stream (particularly at a rate of 200-250ms) have been reported to appear

192 distorted as the relative differences between facial features from one image bleed into the

193 next (Tangen, Murphy, & Thompson, 2011). Given that this effect seems to require holistic

194 processing and is less prominent in inverted faces (Bowden, Whitaker, & Dunn, 2019;

195 Tangen et al., 2011), it is possible that it may distract from our ability to correctly identify

upright faces when presented at certain rates in a rapid stream.

Experiment 2 employs a similar design to Experiment 1; however, to more closely resemble fingerprint identification procedure, participants were shown the target image of a crime scene print first, before viewing the RSVP stream or single comparison print. While this may change the nature of how beneficial the subsequent ensemble representation may be, previous research using the RSVP methodology suggests that when participants are primed to recognise a particular image among a subsequently presented image stream of random images, performance improves drastically (reference), as they now know what to look for. Accordingly, similar to Experiment 1 we predict that performance will improve when viewing more rapid image streams. Additionally, instead of presenting fingerprints in an upright or inverted orientation as in Experiment 1, our conditions manipulated whether participants viewed fingerprints belonging to the same finger (i.e., "Is this John's thumb?"), or to the same person more generally (i.e., "Does this fingerprint belong to John?"), as this manipulation will allow us to simulate the kinds of 'ten-print' materials that fingerprint examiners typically have at their disposal (reference). In doing so, we can examine whether the potential benefits of an RSVP stream are constrained by the specificity of the identification. While evidence suggests that novices may perform similarly when discriminating prints from the same person and same finger (Searston & Tangen, 2017c; Tangen et al., 2011; Tangen et al., 2014), RSVP streams consisting of the same finger prints may contain less variation compared to prints from different fingers from the same person, and therefore may generate a more stable ensemble with which to compare the latent print (see Whitney & Leib, 2018), making recognition easier. We therefore predict that any benefits derived from the RSVP methodology may be more pronounced when viewing streams of the same finger.

In both experiments, while recognition accuracy may be higher when viewing more images, we expect that confidence may be higher when viewing *fewer* images, as these

conditions would likely feel the most intuitive to participants, and would allow participants to maximise the encoding of any particular details.

<div align="center">

Experiment 1

</div>

Methods

The preregistration plan [add link] for this study is available on the Open Science Framework (OSF), and includes our predictions and hypotheses, methodology, power analysis, analysis plan, and links to all available materials, software, raw data files, and R markdown scripts.

Participants. 30 participants took part in this experiment (19 male, 11 female, mean age of 25) consisting of students from the University of Adelaide and members of the general Adelaide population. All participants were required to be at least 18 years of age, fluent in English, and have normal or corrected-to-normal vision. Participants were incentivised by receiving a $20 Coles/Myer gift card in exchange for their time (see Appendix A). All participants provided informed consent prior to commencing the experiment (see Appendix B).

Participants' responses were to be excluded if they failed to complete the experiment due to illness, fatigue or excessive response delays (i.e., longer than the session allows). Participants who responded in less than 500ms, or consecutively provided the same response, for over 30 percent of trials were also to be excluded. In these cases, another participant was to be recruited and given the same stimulus set according to the previous participant's experiment number. None of the 30 participants met any of these pre-specified exclusion criteria.

Design. This experiment had a 4 (presentation rate: single image, 2, 4, 8 images per second) x 2 (orientation: upright vs. inverted) fully within-subjects design. In Experiment 1, participants were presented with a series of 96 face streams for eight seconds. Presentation rate varied across the streams, with participants viewing streams of 64 face images for 125

248 milliseconds each (8 images per second), streams of 32 face images for 250 milliseconds each

249 (4 images per second), streams of 16 images for 500 milliseconds each (2 images per second),

250 and single images of faces for eight seconds. After a brief 500 millisecond delay, a new

251 'target' face image from either the same or different person was displayed and participants

252 indicated on a scale whether they believed this new face was the same or different person as

253 the face in the stream, and their confidence in their decision (see Figure 2).

254   The faces were presented upright for one half of the trials and inverted on the other

255 half. Both orientation blocks were counterbalanced across participants. The four

256 presentation rate blocks were also randomly presented to each participant within the two

257 orientation blocks. Within each presentation rate block, half of the trials depicted the same

258 person as the target image, and the other half depicted a different person to the target

259 image. These trials were randomly presented for each participant.

260   [Figure 2]

261   **Measures.**   Participants indicated their judgments on a 12-point forced choice

262 confidence rating scale: 1 to 6 indicates a "Different" response and 7 to 12 a "Same"

263 response, with ratings closer to 1 and 12 indicating higher confidence than ratings closer to 6

264 or 7 (see Figure 2). This scale allows us to compute participants' accuracy (mean proportion

265 correct), and mean confidence (between 1 and 6), and has been used in previous research to

266 compute individuals' discriminability as indicated by the area under their proper Receiver

267 Operating Characteristic (ROC) curve ('AUC'; Vokey, Tangen, & Cole, 2009).

268   To measure discriminability, we computed each participant's AUC for each condition

269 from their cumulative confidence ratings on same and different trials (see Hanley & McNeil,

270 1982; Vokey, 2016). An AUC of 1 indicates perfect discriminability, and an AUC of .5

271 indicates chance performance. A large number of 'hits' (i.e., participant correctly says

272 "Same") and a small number of 'false alarms' (i.e., participant incorrectly says "Same")

273 indicates high discriminability and would produce an AUC score closer to 1, whereas an

274 equal number of hits and false alarms would indicate chance discriminability, resulting in

275 lower AUC scores closer to .5. Participants' confidence is also taken into account in

276 computing AUC, such that lower confidence judgments reflect lower discriminability.

277   Confidence was computed by collapsing the 12-point rating scale to a 6-point scale.

278 The original scale provided six degrees of confidence for both "Different" (1-6) and "Same"

279 (7-12) responses; and so the collapsed scale isolates confidence by coding all "unsure"

280 responses (6 or 7) to 1, all "moderately unsure" responses (5 or 8) to 2, all "slightly unsure"

281 responses (4 or 9) to 3, and so on—until all "sure" responses (1 or 12) are coded to 6.

282   **Materials.**   The faces were sourced from the VGGFace 2 dataset (Cao, Shen, Xie,

283 Parkhi, & Zisserman, 2018). The original set contains 3.31 million images of 9,131 identities

284 collected from Google Image searches. We used a subset [add link] of 9,600 images of 48

285 identities (200 images per identity). We preserved all natural variation across the images of

286 each identity to increase the difficulty of the target trials (i.e., dissimilar matching identities

287 are more challenging to tell together). The original dataset also contains a large number of

288 blonde, Caucasian, female identities. We constrained our subset to this demographic to

289 increase the difficulty of the distractor trials in the experiment (i.e., similar mismatching

290 identities are more challenging to tell apart). We further increased similarity by computing

291 the distributional characteristics (mean, min, max of image) of each identity and pairing

292 similar identities side-by-side to increase target-distractor resemblance (see Appendix C).

293   We reduced the original set of images for each identity down to 200 by manually

294 excluding any images with dimensions under 100 x 100 pixels, drawings, illustrations or

295 animations of faces, significantly occluded faces, faces with distracting watermarks,

296 duplicates or images that clearly depicted a different identity. All other original details were

297 left intact, including natural variation in pose, age, illumination, etc. We then cropped each

298 face to a square using a script in Adobe Photoshop CC (version 20.0.4) and centred the

299 images around the eyes as close as possible. To avoid ceiling effects for upright faces, we

300 initially reduced all the images to 64 x 64 pixels, then upsized them to 400 x 400 pixels in

301 MATLAB. However, after pilot testing (N = 2) revealed that the task was still too easy for

302 upright faces (mean proportion correct = .92), we further reduced the images to 32 x 32

303 pixels. A second pilot (N = 5) then revealed near-chance performance with the inverted

304 faces (mean proportion correct = .59), and so we generated a fresh batch of images reduced

305 to 48 x 48 pixels to avoid ceiling or chance performance in either condition (see Figure 2).

306 **Software.**   The video instructions and face recognition task were presented to

307 participants on a 13-inch MacBook Pro, with over-ear headphones. We developed the

308 software used to generate the trial sequences, present stimuli to participants, and record

309 their responses in LiveCode (version 9.0.2; the open source 'community edition'). The

310 LiveCode source files and experiment code are available in the Software component of the

311 OSF project. The data analytic scripts and plots for this project were produced in RStudio

312 with the R Markdown package. A list of other package dependencies needed to reproduce

313 our plots and analyses are listed in the data visualisation and analysis html file found in the

314 Analyses component of the OSF project.

315 **Procedure.**   Participants commenced the task after reading an information sheet,

316 signing a consent form, and watching an instructional video [add link]. Participants rated a

317 total of 96 faces as the same or different identity to the faces in the stream. In each case,

318 they indicated their judgments on the 12-point confidence rating scale. The response buttons

319 remained on screen until participants selected their rating; however, a prompt to respond

320 within 4 seconds was displayed between trials if participants took longer to decide.

321 Corrective feedback in the form of an audio (correct or incorrect tone) and visual (the

322 selected response button turns green or red) cue is presented to participants after every trial.

323 The whole face recognition task took about 25 minutes to complete.

Results

\colorbox{yellow}{[Report paired comparisons – and any other instances where significant differences would be unlikely. . .] [Make sure symbols in stats blocks are all correct - generalised eta squared. . .] [insert figures and tables]]

The following analysis examines participants' discriminability (AUC) scores and confidence. Raw proportion correct scores can be found in the Appendix.

Presentation Rate and Orientation.   We conducted repeated measures ANOVAs on participants' AUC scores to test whether their ability to distinguish faces of the same versus different identities significantly increased as presentation rate increased, and whether these effects varied as a function of familiarity with the stimulus orientation. As shown in Table 1, our results suggest that participants are better at recognising faces when viewing rapid streams of the same face compared to single images for both upright and inverted conditions, despite discriminability being lower overall with inverted faces compared to upright faces. A repeated measures ANOVA yielded a significant, medium-to-large (see Cohen, 1988 for conventions) main effect of orientation (F(1, 29) = 68.258, p < .001, G2 = .148) and a significant, small-to-medium main effect of image rate (F(3, 87) = 3.788, p = .013, G2 = .041) on participants' discriminability scores (see Figure 3). No significant interaction was found (F(3, 87) = 1.952, p = .127, G2 = .019). A treatment-control contrast suggested that when compared to viewing a single image, participants' discriminability scores significantly improved under all rapid presentation rate conditions (2 images: t = 2.192, p = .029; 4 images: t = 2.468, p = .014; 8 images: t = 2.431, p = .016). A subsequent trend analysis also revealed a significant linear trend over presentation rate conditions (t = 2.394, p = .018). That is, discriminability increased in a linear fashion as a function of increasing presentation rate for both upright and inverted faces, despite inverted faces being harder to recognise.

Mean Discriminability (AUC)			
Orientation	Image_Rate	mean	SD
upright	1 image	0.548	0.216
upright	2 images	0.715	0.242
upright	4 images	0.698	0.208
upright	8 images	0.684	0.176
inverted	1 image	0.462	0.163
inverted	2 images	0.473	0.202
inverted	4 images	0.513	0.218
inverted	8 images	0.524	0.201

Mean Discriminability (PC)			
Orientation	Image_Rate	mean	SD
upright	1 image	0.619	0.138
upright	2 images	0.733	0.190
upright	4 images	0.733	0.151
upright	8 images	0.733	0.139
inverted	1 image	0.542	0.117
inverted	2 images	0.547	0.145
inverted	4 images	0.625	0.143
inverted	8 images	0.603	0.145

351 [figure 3]

352 To address our prediction that confidence will be highest when viewing single images,

353 we analysed participants' confidence ratings for each condition. As shown in Table 2,

354 participants were more confident at identifying upright compared to inverted faces, though

355 confidence seems similar across different presentation rates. A repeated measures ANOVA

356 revealed a significant, medium-to-large main effect of orientation (F(1, 29) = 8.655, p = .006,

357 G2 = .020), but no significant main effect of image rate (F(3, 87) = 0.785, p = .505, G2 =

358 .002), and no significant interaction (F(3, 87) = 0.365, p = .779, G2 = .001; see Figure 4).

359 Given that confidence did not significantly differ across image rate conditions, our data did

360 not support the third hypothesis.

361 [table 2] [figure 4]

Discussion

363 [Address methodological limitations – e.g., gender, ethnicity, celebrity status] [Address theoretical l

- 364 • Race
- 365 – Other race faces much harder to recognise, presumably because we typically have
- 366 much more experience in recognising faces from our own race
- 367 – Other theories?? Does it exist??
- 368 • Gender
- 369 – ??

370 **Addressing Predictions.**   This experiment aimed to assess what kind of exposure

371 leads to better face recognition when presented with upright and inverted faces. In line with

372 previous face-matching literature (e.g., Murphy et al., 2015; Ritchie & Burton, 2017), our

373 analyses suggest that overall recognition performance increases as participants view more

374 examples of naturally varying face images. This finding builds upon our previous

375 understanding of the ensemble coding literature. While previous research suggests that

376 RSVP streams allow observers to recognise the average representation easier than individual

377 instances in the stream (e.g., Ariely, 2001; De Fockert & Wolfenstein, 2009), the current

378 study suggests that this ensemble can also facilitate the recognition of new instances of the

379 same category. This is not surprising, given that previous face recognition research suggests

380 that we compare new instances of a familiar face to the average representation of that face in

381 our long-term memory (e.g., Bruce & Young, 1986; Burton & Bruce, 1993).

382   Our results also suggest that the benefit associated with increasing image rate occurred

383 in a similar manner for both upright and inverted faces, despite inverted faces being harder

384 to recognise overall. While lower performance when recognising inverted faces was expected

385 (see Tanaka & Simonyi, 2016, and Valentine, 1988), it is surprising that the RSVP paradigm

386 influenced both upright and inverted faces equally. Given that we already process upright

387 faces more successfully than inverted faces, possibly due to experience (Tanaka & Simonyi,

388 2016), we expected that image streams may only provide a slight benefit over single images,

389 compared to inverted faces, which may show a larger benefit as image rate increased. The

390 fact that the two orientation conditions increased in a similar manner may be a product of

391 presenting the images at a reduced resolution. During pilot testing, we blurred the images to

392 increase difficulty with upright faces and prevent ceiling effects (e.g., Balas, Gable, &

393 Pearson, 2019). It is possible, therefore, that while an advantage for upright face processing

394 is still evident, it may be less prominent at low resolutions, allowing the image streams to

395 demonstrate a similar advantage for both orientation conditions. However, no studies seem

396 to have tested the face inversion effect at reduced resolutions, and so future research may

397 wish to confirm this conclusion.

398   I also suspected a lesser advantage for upright faces due to the flashed-face distortion

399 effect (FFDE). The FFDE refers to the apparent distortion of upright (but not inverted)

400 faces presented in an RSVP stream of different random faces, and is thought to emerge due

401 to the relative differences between facial features contrasting from one identity to the next

402 (Tangen et al., 2011). The lack of interaction between orientation conditions, however,

403 suggests that the FFDE had no detrimental effect on either condition. Given that each face

404 in the streams belonged to the same person in the current experiment, rather than different

405 people as is typically the case with FFDE studies (e.g., Balas & Pearson, 2019; Bowden et

406 al., 2019), it may be that the commonalities across each face image were exaggerated, rather

than the differences, thereby increasing performance when viewing rapid streams. However, given that I did not directly manipulate the FFDE, future experiments may wish to explicitly measure the potential influence of this effect in similar face recognition tasks, to investigate whether it aids encoding of an unfamiliar face.

Limitations.   One minor limitation regarding the current methodology is that, given that the selected database sampled faces from Google Images, several of the identities depicted celebrities. Although this provided a suitably large sample of naturally varying face images that could not be found in other databases, this may have increased participants' performance in some trials and inflated my effect sizes, as familiar faces are easier to recognise than unfamiliar faces (Megreya & Burton, 2006). Although an informal post-experiment assessment of each participant's prior familiarity with each face demonstrated that most participants were unfamiliar with most of the faces regardless, future research may wish to use a dataset containing exclusively unfamiliar faces if possible.

Experiment 2

Experiment 1 suggests that presenting similar images in an RSVP stream can facilitate the identification of new instances even when viewing less familiar stimuli (e.g., inverted faces). This method of rapidly presenting multiple similar instances may also be useful in improving performance in other disciplines that rely on identifying naturally varying images—such as fingerprint examination (see Figure 5).

Method

In this experiment, participants viewed single images of a latent crime scene fingerprint before viewing a stream of fingerprint images. They then determined whether the fingerprints in the stream belonged to the same or different finger, or the same or different person more broadly, to the latent fingerprint (see Figure 5 and Figure 6). As in Experiment 1, presentation rate varied for each stream, and participants' confidence and discriminability

were the main performance measures of interest. This experiment was preregistered along with Experiment 1.

Participants.   Both experiments were conducted concurrently with the same participants.

Design.   Experiment 2 had a 4 (image presentation rate: single image, 2, 4, 8 images per 8-second stream) x 2 (image specificity: prints from the same finger vs. prints from the same person) fully within-subjects design. Participants judged if a latent fingerprint belonged to the same or different finger or person as the fingerprint images in a rapidly presented stream of images. In this experiment, participants viewed the latent fingerprint (single image) before viewing the image stream. Due to the limited number of fingerprint images in the selected dataset, streams consisted of one-second fingerprint streams presented 'on loop' for eight seconds. Participants viewed streams of eight images per second for 125 milliseconds each, streams of four images per second for 250 milliseconds each, streams of two images per second for 500 milliseconds each, and single fingerprint images for eight seconds. Fingerprint streams remained on-screen until a response was made, though participants were prompted to respond within eight seconds (see Figure 6). Participants received corrective feedback for every decision.

[figure 6]

Materials.   The fingerprints were generated from a subset of the Forensic Informatics Biometric Repository (Tear, Thompson, & Tangen, 2010). For the person recognition component of the task, there are ten fully-rolled prints, one from each finger, from 48 different individuals. These served as the rolled prints presented in the rapid streams. For each individual there is also one 'target' latent print from the same person, and a 'distractor' latent print from another person. The targets and distractors were always taken from the left thumb, as previous research suggests that novices can distinguish prints based on hand type (less so based on finger type; Searston & Tangen, 2017a, 2017b; Thompson & Tangen, 2014). For the finger recognition component of the task, there are eight different fully-rolled

impressions from the left thumb of the same 48 individuals. The target and distractor latent

prints are the same as those used in the person component of the task.

All natural variation in the latent prints was preserved, while the rolled prints

presented in the streams were centred on a white background, grey-scaled, level balanced,

and cropped to 400 x 400 pixels (as with the faces). Any distracting borders and text from

the arrest cards were removed to isolate the prints.

Software.   The software for Experiment 2 was identical to that in Experiment 1. The

relevant files are similarly available under the same pre-registration link.

Procedure.   Participants were randomly assigned to complete Experiment 2 either

immediately before or after Experiment 1. The procedure for Experiment 2 was identical to

that in Experiment 1, except for the necessary design changes, and participants were

prompted to respond within eight seconds.

Results

[Report paired comparisons – and any other instances where significant differences would be unlike]

The following analysis examines participants' discriminability (AUC) scores and confidence.

Raw proportion correct scores can be found in the Appendix.

Presentation Rate and Image Specificity.   I conducted repeated measures

ANOVAs on participants' AUC scores to test whether their ability to distinguish related and

non-related fingerprints significantly increased as presentation rate increased, and whether

these effects varied as a function of stimulus specificity level. As shown in Table 3, my results

show that participants' fingerprint recognition performance generally decreased as image rate

increased for both "same finger" and "same person" conditions. My results suggest no

significant main effect of specificity ($F(1, 29) = 0.108$, $p = .744$, G2 < .001), a significant,

small-to-moderate main effect of image rate ($F(3, 87) = 3.367$, $p = .022$, G2 = .035) on

participants' discriminability, and no significant interaction ($F(3, 87) = 2.053$, $p = .112$, G2

= .018; see Figure 7). Mauchly's test for sphericity suggests that the assumption of sphericity

was met (image rate: W = .934, p = .862; specificity-image rate interaction: W = .827, p = .386); and so no corrections were applied to the reported p-values. A treatment-control contrast suggested that compared to viewing a single image, participants' discriminability scores significantly decreased when presented with 4 and 8 images per second (2 images: t = -0.897, p = .371; 4 images: t = -2.016, p = .045; 8 images: t = -2.663, p = .008). A subsequent trend analysis also revealed a significant linear trend over presentation rate (t = -2.880; p = .004). That is, discriminability decreased in a linear fashion as presentation rate increased for both same finger and same person conditions—contrary to my predictions.

[table 3] [figure 7]

To investigate my prediction that confidence will be highest when viewing single images, I also examined participants' confidence ratings for each condition. As demonstrated in Table 4, participants were consistently confident across all presentation rates when viewing streams of prints from the same person and prints from the same finger. A repeated measures ANOVA revealed no significant main effect of specificity (F(1,29) = 3.994, p = .055, G2 = .006) or image rate (F(3,87) = 0.763, p = .518, G2 = .002), and no significant interaction (F(3,87) = 0.486, p = .693, G2 < .001; see Figure 8). Mauchly's test for sphericity suggests that the assumption of sphericity was met (image rate: W = .743, p = .144; specificity-image rate interaction: W=.676, p = .054); and so no corrections were applied to the reported p-values. Given that confidence did not significantly differ across image rate conditions, my data does not support my initial prediction.

[table 4] [figure 8]

Discussion

Addressing Predictions. This experiment aimed to assess whether viewing several impressions of similar fingerprints, either from the same finger or the same person, would better assist novices in making an identification compared to viewing a single fingerprint for

a longer duration. My results suggest that this is not the case for either condition. Since novices have no experience in fingerprint matching, it is possible that recognition may benefit from carefully assessing fingerprints, as is currently standard practice (e.g., Busey & Parada, 2010), during the early stages of training. Indeed, given that understanding the images in an RSVP stream seems to rely on holistically processing each image (i.e., perceiving a complex image as a whole, rather than a collection of features; see Oliva, 2005), which may depend on image familiarity (e.g., Tanaka & Simonyi, 2016), it may be that the completely novel nature of the stimulus class required longer exposure to compensate for a lack of holistic processing. If this is true, it is plausible that rapidly presenting fingerprints may have introduced a floor effect in participants' performance—obscuring any positive effect that viewing multiple exemplars may have otherwise exerted. This explanation seems likely, as discrimination performance significantly decreased as presentation rate dropped below 300 milliseconds per image—the approximated minimum duration required to process visual stimuli (Potter, 1976).

The fact that there was no significant difference or interaction between the same person and same finger conditions was also surprising. I suspected that performance would be higher when participants viewed streams from the same finger, to the extent that these streams contain less variation compared those in the 'same person' condition, thus providing a more stable ensemble representation with which to compare the latent print (see Whitney & Leib, 2018). However, while no studies have directly compared the two conditions as in the present experiment, evidence suggests that novices may not perform very differently when asked to match a print to either the same person or same finger (see Searston & Tangen, 2017c, Tangen et al., 2011, and Thompson et al., 2014). It seems likely, therefore, that because novices have no specific fingerprint matching experience like experts, the RSVP methodology allows them to notice general similarities between related prints, regardless of how precisely the prints are related.

Future Directions. While the current results suggest that the RSVP paradigm does not improve fingerprint novice performance, this does not necessarily mean that exposure to various naturally varying fingerprints will not benefit novices. Previous research suggests that images presented in streams of at least one second per image can be efficiently remembered for long periods (e.g., Potter & Levy, 1969; Standing, 1973); and additionally, Thompson and Tangen (2014, Experiment 3) suggested that viewing a print for two seconds only incurred a 6.8 percent decrease in accuracy for novices compared to viewing prints for one minute. It is possible, therefore, that if each fingerprint in the stream was presented for several seconds, rather than several milliseconds, this may optimally balance the advantages of both viewing the detail in a single image and being exposed to variability within images. Future research may wish to either decrease the presentation rate, or allow participants themselves to control presentation rate and view each fingerprint for as long as they deem necessary for familiarisation. The latter manipulation would preserve individual differences in evidence accumulation styles (i.e., some people may prefer more image variation, while others may prefer more viewing time), providing a less intrusive method of investigating how presentation rate might predict identification performance. Additionally, future research may wish to administer the current experiment to participants with varying degrees of fingerprint-matching experience. Given that novices did not benefit from the RSVP stream (and were no better than chance in some conditions), it is possible that more experienced fingerprint examiners may derive greater benefits from the RSVP paradigm, as they may process the fingerprints more holistically (Busey & Vanderkolk, 2005; but see Vogelsang, Palmeri, & Busey, 2017 for a competing study). Given that previous research suggests that the majority of learning among novices occurs within the first three months of training (Searston & Tangen, 2017b), it is possible that increasing exposure to varying prints may be most beneficial after the initial learning phase.

General Discussion

This thesis examined whether rapidly viewing several instances of complex stimuli, across varying levels of familiarity (Experiment 1) and specificity (Experiment 2), would better facilitate recognition of a new instance compared to viewing a single image for a longer duration. Previous literature suggests that we can recognise new instances of an object based on our prior experience with similar instances (Brooks, 1987; Medin & Ross, 1989). Research on ensemble coding also suggests that we can rapidly understand the general nature of an object as we view several similar, varying instances (e.g., Im & Chong, 2009; Morgan et al., 2000). However, no research has examined how an RSVP-generated ensemble representation may assist in identifying new instances.

Experiment 1 suggests that ensemble coding may indeed facilitate recognition when viewing upright and inverted faces. Given that upright and inverted faces differ only in observers' decreased familiarity with inverted faces (Valentine, 1988), these results suggest that ensemble coding may assist recognition even when exposed to less familiar stimuli. Experiment 2, however, suggests the opposite pattern of results, as fingerprints—a completely unfamiliar stimulus class—showed worse discrimination when participants were presented with RSVP streams from either the same finger or same person as the crime scene print.

Addressing Predictions.   Contrary to my predictions in both experiments, participants' confidence showed no significant differences across image rate conditions, despite single images allowing the greatest encoding time. It may be that the task demands were too difficult in each condition for participants to feel confident. Indeed, identifying different instances of unfamiliar faces has been reported to be a challenging task (e.g., Bruce et al., 1999), which would undoubtedly be harder when the faces are blurred (e.g., Balas et al., 2019; Sanford, Sarker, & Bernier, 2018); and novice performance in fingerprint matching appears equally challenging (Searston & Tangen, 2017c; Tangen et al., 2011; Thompson et al., 2014). It seems likely that the relative disadvantages in either condition (i.e., less

587 variation with single images compared to less processing time with several images) may have

588 undermined confidence equally across all conditions.

589   **Discrepancies Between Discriminability Patterns.**   Although my contradicting

590 discriminability results between the two experiments were unexpected, several explanations

591 are possible. Firstly, the fact that I presented the test stimulus in Experiment 2 before,

592 rather than after the image streams, may have placed greater demands on working

593 memory—especially as the 'more familiar' faces in Experiment 1 (approximated from rapid

594 stream conditions) may have already demanded less from working memory compared to

595 recognising 'less familiar' faces (approximated by single image conditions; Jackson &

596 Raymond, 2008). As opposed to Experiment 1, where the test stimulus remained onscreen

597 until the response, participants in Experiment 2 had to hold a complex, unfamiliar, noisy

598 latent fingerprint in working memory while viewing the subsequent print streams. This

599 working memory demand may have made Experiment 2 more difficult than Experiment 1,

600 particularly as the images became more difficult to process at faster image rates. The fact

601 that ensemble coding seems more beneficial during the encoding stage of learning an identity,

602 rather than on retrieval, seems concurrent with previous research on categorisation. Such

603 research typically suggests that we can identify a new image by comparing its similarity to

604 previously encountered images or representations (e.g., Brooks, 1987; Dopkins & Gleason,

605 1997). If participants can only view similar instances after being exposed to the test

606 stimulus, as in Experiment 2, then they are not previously encountering similar instances to

607 create a representation; they view these images after the fact.

608   A second possible explanation is that compared to upright and inverted faces,

609 fingerprints may be too difficult for novices to process using the current methodology.

610 Although Experiment 1 suggests that RSVP streams may familiarise observers with less

611 familiar stimuli, fingerprints may simply be too unfamiliar for a similar benefit to occur. The

612 RSVP methodology seems to depend on holistic processing (see Oliva, 2005), and while

613 previous research suggests that we process unfamiliar stimuli less holistically than familiar

614 stimuli (e.g., Campbell & Tanaka, 2018; Wong et al., 2009), holistic and analytic processing

615 seem to be opposing ends of a spectrum, rather than a dichotomy (see Farah, 1992, and

616 Tanaka & Simonyi, 2016). That is, while inverted faces are not processed as holistically as

617 upright faces (Tanaka & Simonyi, 2016), fingerprints may be processed even less so, and

618 therefore benefit less from the RSVP paradigm as presentation rate increases. Future

619 research may wish to confirm these suspicions, assessing and comparing our holistic

620 processing abilities with a range of less familiar stimuli (e.g., fingerprints, paintings, bird

621 species) with various recognition or categorisation tasks.

622 **Discrepancies Between Chance Comparisons.** While participants in both

623 experiments displayed better performance than chance, participants in Experiment 1

624 displayed a higher difference (d = 0.121) than those in Experiment 2 (d = 0.058). In

625 addition to the changes listed above, this difference in overall discriminability may be due to

626 the fact that Experiment 1 had a higher degree of image variation than Experiment 2. In

627 Experiment 1, all images were coloured and blurred and consisted of people in different

628 contexts, including the subsequent test images; however, in Experiment 2 the stream images

629 were somewhat controlled and artificial (i.e., fully-rolled prints, all on a white background)

630 compared to the latent crime scene prints, which may vary in different ways to the prints

631 used in the stream (e.g., contact surface or print pressure). That is, the streams in

632 Experiment 1 were a closer match to the test images than in Experiment 2. Previous

633 research in face recognition suggests that exposure to more variable images better facilitates

634 recognition in a new context compared to less variable images (Menon, White, & Kemp,

635 2015; Ritchie & Burton, 2017), and so it is possible that the more controlled nature of the

636 stream images in Experiment 2 may have hindered participants' ability to recognise the test

637 images compared to the more variable stream images in Experiment 1. However, Ritchie and

638 Burton (2017) suggest that [viewing multiple similar images, even with (?)] reduced

639 variability should nevertheless increase rather than decrease recognition compared to viewing

640 single images. As such, while reduced variability may explain why participants did not

benefit from the print streams in Experiment 2, it does not account for the significant

decrease in discriminability observed with increasing presentation rates. Of course, it is

possible that a combination of the aforementioned design factors may have produced the

opposite trends observed across the two experiments.

Another possible factor that may have contributed to the different pattern of results

across the two experiments is that Experiment 2 contained fewer unique image exemplars in

the streams compared to those in Experiment 1. Given the differences in the selected

databases, participants viewed fewer unique fingerprints in each stream compared to the

faces in Experiment 1. Indeed, even the highest presentation rate condition in Experiment 2

only showed participants eight unique prints, compared to the slowest stream condition in

Experiment 1, which contained 16 unique faces. Given that previous research suggests that

viewing fewer different exemplars may decrease recognition of new instances compared to

viewing more (Murphy et al., 2015), it is possible that there were not enough fingerprints to

produce a similar benefit of presentation rate in Experiment 2. However, it is also important

to note that, in real-world fingerprint examination settings, examiners are unlikely to always

have access to many varying exemplars of a suspects' fingerprints—in some cases, fingerprint

databases may only contain a single comparison print, or a ten-print card consisting of

fully-rolled prints and 'slapped' prints from the same person, and not the same finger (Jain,

Nandakumar, & Ross, in press; PCAST, 2016). While Experiment 2 aimed to use prints that

fingerprint analysts are likely to encounter in their daily work (e.g., latent crime scene prints

presented with fully rolled suspect prints), and the aforementioned task constraints are an

important limitation with respect to the experiment's theoretical implications, they also

highlight real constraints in attempting to generalise these findings to more applied contexts.

Broader Implications.   Despite the different pattern of results observed with faces

and fingerprints, my findings nevertheless help reveal important information about how

observers may best familiarise themselves with novel images under different conditions. If

these findings were to be replicated or extended in different contexts, they may reveal

benefits of image presentation rate beyond face recognition for other domains of perceptual expertise. Given that prior exposure to variation seems to increase recognition performance when controlling for time, the identification decisions of counterfeit investigators, passport officers, various medical practitioners, and other professionals who rely on their perceptual expertise, may benefit from accumulating as much exposure as possible to varying examples within their domain. Future research may look to improve expert identification decisions by optimising the advantages of viewing time and exposure to variation in a range of given fields.

- Experts (e.g., fingerprints, antique cars) struggle to identify things too far from their domain of expertise... possible that exp 2 will yield different results depending on whether we test experts or not

Conclusion

This thesis is the first to explore how to best familiarise observers with complex, unfamiliar images given a fixed amount of time: should we assess the finer details, or glean the general gist of several similar images? Across two experiments, I establish a new relationship between the RSVP-based ensemble coding literature and the image recognition literature, with the caveat that this relationship may change when presented under different conditions and in other expert domains not explored in this thesis. In Experiment 2, I attempted to boost novices' fingerprint identification performance by increasing their exposure to fingerprint variation in each case, and I found tentative support for current analytical practices, as reported by analysts, during the early stages of their training. My thesis highlights the need to further investigate how to optimally balance the potential advantages of both assessing the details of individual instances, and gaining experience with natural variation, when tasked with recognising familiar or unfamiliar identities and visual categories. As it stands, this thesis provides foundational evidence for the effect of presentation rate that may inform future research on improving the training and

694 identification decisions of professionals in medicine, security, and law enforcement—who are

695 faced with the task of diagnosing or classifying new complex cases based on their previous

696 experience.

References

697

```
r_refs(file = "r-references.bib")
```