Cognition 186 (2019) 42–49

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Decisional space modulates visual categorization – Evidence from saccadic reaction times

Meike Ramon[1,*], Nayla Sokhn[1], Roberto Caldara

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Switzerland

ARTICLE INFO

Keywords:
Face processing
Visual categorization
Minimum saccadic reaction times
Personal familiarity
Face gender
Task demands

ABSTRACT

Manual and saccadic reaction times (SRTs) have been used to determine the minimum time required for different types of visual categorizations. Such studies have demonstrated extremely rapid detection of faces within natural scenes, whereas increasingly complex decisions (i.e. levels of processing) require longer processing times. We reasoned that visual categorization speed is not only dependent on the processing level, but is further affected by decisional space constraints. In the context of two different tasks, observers performed choice saccades towards female (gender categorization) or personally familiar (familiarity categorization) faces. Additionally, familiarity categorizations were completed with stimulus sets that differed in the number of individuals presented (3 vs. 7 identities) to investigate the effect of decisional space constraints. We observe an inverse relationship between visual categorization proficiency and decisional space. Observers were most accurate for categorization of gender, which could be achieved in as little as 140 ms. Categorization of highly predictable targets was more error-prone and required an additional ~100 ms processing time. Our findings add to an increasing body of evidence demonstraing that pre-activation of identity-information can modulate early visual processing in a top-down manner. They also emphasize the importance of considering procedural aspects, as well as terminology when aiming to characterize cognitive processes.

1. Introduction

Across sensory modalities, stimulus categorization can involve a number of different processes. Different types of stimuli can be detected, discriminated, recognized as having been encountered before, or identified on the exemplar level. The human visual system can perform such categorizations with high proficiency despite large variations in stimulus input. For instance, categorization of animals presented in images of natural scenes presented for as little as 20 ms can be reliably achieved in under 300 ms (ms) (e.g., Macé, Thorpe, & Fabre-Thorpe, 2005; Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe, Fize, & Marlot, 1996). The speed of such categorical responses can be exploited to study the characteristics of visual processing. Specifically, the time required to perform accurate visual categorizations provides a valuable source of information, which can be levered to constrain theories of visual processing (for a review see e.g., Fabre-Thorpe, 2011). However, a number of other aspects can influence visual categorization speed.

One important aspect that affects visual categorization speed is the effector from which responses are recorded, with certain responses requiring more time to be executed than others. Estimates of visual categorizations have been derived from verification and naming tasks, which typically involve longer reaction times (RTs) (Tanaka & Taylor, 1991), as compared to manual forced choice paradigms, where subjects press (or release) one of two buttons (a button) to distinguish between stimulus categories (e.g., Hasbroucq, Mouret, Seal, & Akamatsu, 1995). To reduce response times, more sensitive Go/no-go paradigms have been developed that require subjects to respond to predefined target categories via finger-lift from an infra-red sensor (e.g., Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Rousselet et al., 2002; Thorpe et al., 1996; VanRullen & Thorpe, 2001). Finally, other paradigms exploiting the very rapid responses of oculo-motor effectors have been developed. The saccadic reaction times (SRTs) measured in such paradigms provide a more precise description of the lower bound of speeded categorizations (Crouzet, Joubert, Thorpe, & Fabre-Thorpe, 2009; Crouzet, Kirchner, & Thorpe, 2010; Kirchner & Thorpe, 2006). Together, these developments have served the goal of maximizing the precision estimates of visual processing speed (for a direct comparison of paradigms, see Bacon-Macé, Kirchner, Fabre-Thorpe, & Thorpe, 2007). Moreover, aside from deriving mean or median RTs, these paradigms have aimed to determine the *minimum* reaction time

* Corresponding author at: University of Fribourg, Department of Psychology, Faucigny 2, 1700 Fribourg, Switzerland.
E-mail address: meike.ramon@gmail.com (M. Ramon).

[1] Both authors contributed equally to this work

https://doi.org/10.1016/j.cognition.2019.01.019
Received 29 March 2018; Received in revised form 24 January 2019; Accepted 29 January 2019

(minRT). The minRT is defined as the first time-bin for which correct responses significantly outnumber incorrect ones (Fabre-Thorpe, Richard, & Thorpe, 1998; VanRullen & Thorpe, 2001), and is considered to reflect the minimal processing time required for *reliable* responses (Rousselet, Macé, & Fabre-Thorpe, 2003).

A second aspect that affects visual categorization speed is the nature of the stimulus processed. In the field of visual cognition, numerous studies have investigated the role of stimulus category on categorization proficiency. Such studies have e.g. demonstrated that human faces can be rapidly detected within natural scenes within as little as 100–110 ms (Crouzet et al., 2010), and are more efficiently processed compared to other non-face objects (e.g., Caudek, 2013; for a review see Yovel, 2016). These findings have supported the notion that faces represent a "special" stimulus category, studied by a large and continuously growing body of research, which can generate insights into general functioning principles of the human brain.

A third aspect that affects visual processing speed is the process engaged in, or processing level (cf. Mack & Palmeri, 2015). Faces are ideal stimuli to study visual categorization, because they can be processed at multiple levels. For example, typical observers can accurately and rapidly detect the presence of a face, determine its gender, age, and ethnicity, discriminate between and recognize previously encountered persons, and finally identify unique individuals and recall associated semantic information. Psychophysical studies have demonstrated that different visual categorization processes are associated with important differences in information diagnosticity. For instance, the ability to detect a face, determine its gender, expression, or familiarity, relies on different facial information (Schyns, 1998; Gosselin & Schyns, 2001; Schyns, Bonnar & Gosselin, 2002; Smith, Fries, Gosselin, Goebel & Schyns, 2009; Smith, Volna & Ewing, 2016). Importantly, these different categorization processes are associated with varied levels of behavioral proficiency. For example, superordinate human vs. animal decisions can be performed manually with high precision (98%) in as little as 285 ms, while famous vs. unfamiliar manual decisions are more error-prone and slower (75%, 468 ms); gender categorization, on the other hand, can be performed with high fidelity and intermediate minRTs (94%; ~310 ms) (Barragan-Jason, Lachat, & Barbeau, 2012). Adding to these findings, we recently reported processing level dependent categorization proficiency for both healthy observers and a case of acquired prosopagnosia (Ramon, Sokhn, Lao & Caldara, 2018).

Despite such evidence, the majority of studies of visual categorization do not systematically address or consider the impact of different processing levels. Additionally, and perhaps consequently, many studies lack terminological precision. For instance, some researchers use the term "face identification" for any task that deals with processing of facial identity, e.g. face matching (e.g., Balsdon, Summersby, Kemo & White, 2018, who also use the term "face recognition" interchangeably). Similarly, while "face recognition" refers to the ability to identify stimuli as previously learned (e.g., Duchaine & Nakayama, 2006; Blais, Jack, Scheepers, Fiset & Caldara, 2008; Barton & Corrow, 2016; Ramon & Gobbini, 2018), others misleadingly use this term in the context of face matching tasks (e.g., Phillips, Yates, Hu, Hahn, Noyes, et al., 2018). Both indiscriminate terminology, as well as lack of consideration of process(es) implied can generate seemingly contradictory findings, which can be detrimental to advancing our understanding of visual categorization (Ramon, 2018; Ramon, Bobak & White, in press).

However, type of effector, stimulus category, and processing level are not the only factors that flexibly modulate visual categorization on the neural and behavioral level (cf. Mack & Palmeri, 2015; Praβ, Grimsen König & Fahle, 2013; Kim & McCarthy, 2016; Rotshtein, Schfield, Funes & Humphreys, 2010; Bonte, Hausfeld, Scharke, Valente & Formisano, 2014). Additionally, observers' experience, environmental regularities, and expectations determine processing proficiency (e.g., Palmeri & Mack, 2015; Hall, Mattingley & Dux, 2018; Manahova, Mostert, Schoffelen & de Lange, 2018; Eger, Henson, Driver & Dolan, 2007; Sauvé, Harmand, Vanni & Brodeur, 2017; Kadel, Feldmann-Wüstefeld & Schubö, 2017; for reviews see Summerfield & Egner, 2009; Lupyan, 2015; O'Callaghan, Kveraga, Shine, Adams & Bar, 2017; Seriès & Seitz, 2013). A simple way to demonstrate this is by varying the size of the stimulus set used, which (together with other procedural choices) determines an item's probability of being encountered.

A large body of work has demonstrated the important effect that stimulus set size can have on neural processing and overt behavior. Several studies support the idea that higher (neural and behavioral) search efficiency is found for smaller categories (cf. Wu, McGee, Echiverri & Zinszer, 2018). Psychophysical studies involving target search for simple stimuli in multi-item displays have demonstrated that visual search proficiency declines as the number of simultaneously available distractors increases (e.g., Palmer, 1994; Busey & Palmer, 2008). Such decreases in performance accuracy and RT increases observed with increasing set sizes, referred to as set size effects, are found across sensory modalities, as well as for individually presented stimuli of varied complexity (e.g., Pollack, 1952; Kent & Lamberts, 2005; Lacouture & Marley, 2004). Concerning high-level vision, as investigated by face processing, set size (or attentional load; cf. Olk & Garay-Vado, 2011) has been reported to affect visual categorization. For example, interactions between invariant features (e.g., emotional expression, gender, race) varies as a function of set size (Lipp, Karnadewi, Craig & Cronin, 2015; Craig, Mallan & Lipp, 2012). Crucially, set size has been reported to account for "apparent inconsistency in the literature on face categorization" (Lipp et al., 2015; p.1293).

As such, two experiments can involve the same stimulus category (e.g., faces) and task (e.g., pressing a button when an exemplar of a target category is presented). Independent of these similarities, both can involve distinct processes (e.g., detecting the occurrence of a female face, or that of a familiar individual), for which potentially different sources of information are diagnostic, which are in turn associated with different levels of categorization proficiency. Performance will also be determined by two additional factors: the probability of seeing an exemplar from a target category (i.e., the proportion of female to male, and unfamiliar to familiar faces); the total number of exemplars per category, i.e. stimulus set size. Specifically, if individual level processing is required (e.g., in familiarity decisions and identification tasks), small set sizes should be associated with more proficient performance. Simply speaking, it is easier to determine whether an exemplar is one of few, as opposed to one of many.

Therefore, studying visual categorization requires careful consideration of a number of aspects: effector type, stimulus category (accurate definition and terminology related to), processing level, and stimulus predictability, which is partly determined by the stimulus set size. Neglecting any of these aspects can promote erroneous conclusions and apparently discrepant findings. To illustrate, consider recent findings of process-dependent face categorization. Numerous findings suggest that face detection precedes familiarity recognition. For example, Besson et al. (2017) assessed face categorization at three different levels through manual minRTs using extremely large stimulus set to avoid image repetitions. The authors reported highly accurate performance and very rapid minRTs for "human face categorization" (i.e., face detection ~240 ms) and "individual face recognition" (i.e., searching for a single predefined target identity; ~260 ms), but more error prone and slower responses for "familiar face recognition" (i.e, deciding whether a face belonged to a large pool of famous individuals; ~380 ms), as also reported for personally familiar face recognition (Ramon, Caharel, & Rossion, 2011). Thus, 380 ms represents the lower bound for speeded manual familiarity decisions that occur ca. 180 ms after the earliest neural marker of familiarity (e.g., Barragan-Jason, Cauchoix, & Barbeau, 2015; Caharel, Ramon, & Rossion, 2014; Huang et al., 2017). Seemingly contradictory evidence was reported by Visconti di Oleggio Castello and Gobbini (2015). Measuring the minimum speed of choice saccades expressed towards personally familiar (PF) faces, the authors concluded that personally familiar face "detection" could be achieved in 180 ms, before "explicit recognition of identity" (p.1). These findings

Fig. 1. Examples of photographs of personally familiar and unfamiliar stimuli taken for stimulus creation.

are incompatible with electrophysiological studies, which have reported the earliest familiarity-dependent differential response to faces (~140–200 ms; e.g. Barragan-Jason et al., 2015; Caharel et al., 2014; Huang et al., 2017).

In our opinion, the most parsimonious account for these apparently discrepant findings lies in procedural differences and misleading terminology (cf. Ramon, 2018; Ramon et al., 2018). Visconti di Oleggio Castello and Gobbini's (2015) observers performed choice saccades towards a set of familiar faces comprising merely three target identities (which differed across observers). Thus, given this highly constrained *decisional space*, the process that Visconti di Oleggio Castello and Gobbini (2015) measured is much less "detection" (i.e. spotting one of numerous possible instances of a category). Rather, it represents a three-alternative "individual face recognition" (Besson et al., 2017) task, which can be achieved in < 140 ms for a single target identity (Mathey et al., 2012). Despite repeating each image several hundreds of times, Visconti di Oleggio Castello and Gobbini's (2015) observers' performance in the three-alternative target search task was inferior to that reported by Mathey et al. (2012) in their *single* target task (performance ranges: 49–69% vs. 60–75%).

Here, we sought to systematically investigate the extent to which visual categorization speed is affected by two factors: processing level and decisional space. To this end, we employed a time-sensitive SRT paradigm, in the context of two different processes performed on a single stimulus category (i.e., faces): gender and familiarity decisions expressed through two alternative forced-choice saccades (cf. Ramon et al., 2018). Additionally, in the context of familiarity decisions, we manipulated the decisional space by varying the number of target exemplars, thereby creating a 1-of-*n*-target search task. This effectively manipulates observers' expectation given the varied probability of exemplar presentation: both familiarity decision tasks differed in that their decisional space was either comparatively broad or narrow.

To anticipate our findings, in line with previous work (Ramon et al., 2018; Mathey et al., 2012) rapid minSRTs are observed under conditions of high predictability, where the search space is confined to a binary, unambiguous category (gender decision task), or a very small number of target items (familiarity decision with few identities). These findings indicate that "detection of personal familiarity" requires significantly more time than previously reported, unless the term "detection" is used loosely to describe responses executed towards an extremely restricted number of target identities (Visconti di Oleggio Castello & Gobbini, 2015) – a process that should more accurately be labeled as n-alternative target search (Besson et al., 2017).

2. Methods

2.1. Participants

We tested three groups of subjects: the first ($n_{lab\_members} = 8$; 4 females, mean age: 31, range: 24–37 comprised group members from the iBMLab. Group members were highly familiar with the target identities, who were their colleagues for several years; this first group completed all three tasks described below. The second and third groups ($n_{students\_gender} = 36$; 31 females, mean age: 21, range: 20–24; $n_{students\_familiarity} = 27$; 24 females, mean age: 22, range: 20–24) comprised students from the Department of Psychology who knew the members of Department depicted in the stimulus material through their teaching and mentoring activities. The second group of observers completed the gender categorization task; the third group completed both personal familiarity categorization tasks. Note that observers of the second and third groups completed only one paradigm (gender or familiarity decisions), due to internal departmental regulations regarding the number of experimental hours completed by student participants. Groups' performance (i.e., lab members vs. students) was considered separately in the analyses due to the inherent differences in age and exposure to individuals whose likenesses were presented in the experiments. All participants provided written informed consent; all procedures were approved by the internal ethics committee of the Department of Psychology at the University of Fribourg, Switzerland and are in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2. Stimuli

The full stimulus set comprised natural (uncropped, color) images of 14 facial identities (7 un/familiar) taken from three different viewpoints (frontal, left, right). For each PF identity, images of a corresponding unfamiliar identity carefully matched for age, gender, and appearance (hair color and style, eye color) were taken (see Fig. 1 for stimulus examples). Image processing included placement on a uniform grey background (630 × 630 pixels) and correction for low-level properties (luminance, contrast) using the SHINE toolbox (Willenbockel et al., 2010), as well as additional ones kindly provided by V. Willenbockel to allow for equation of color stimuli.

2.3. Procedures

Prior to completing the experiments, subjects completed familiarity ratings to determine their level of familiarity with each identity (PF and unfamiliar) presented. Each item of the familiarity questionnaire consisted of an image of each stimulus identity taken under varied, natural conditions. Lab members' images were taken from professional websites, unfamiliar identities' images were taken from social media. Observers had to indicate their self-reported degree of familiarity with each individual on a scale from 1 (not at all familiar) to 5 (highly familiar). For all experiments, stimuli were presented on a 1920 × 1080 pixel VIEWPixx monitor. Subjects' oculo-motor behavior was recorded at a sampling rate of 1000 Hz with an SR Research Desktop-Mount EyeLink 2 K eye tracker (with a chin and forehead rest; average gaze position error ~0.5, spatial resolution: ~0.01). The eye-tracker had a linear output over the range of the monitor used. Although viewing was binocular, only the left eye was tracked; given the fully balanced stimulus presentation across visual fields, inter-individual differences in ocular dominance were considered irrelevant. The experiment was implemented in Matlab (R2009b, The MathWorks, Natick, MA), using the Psychophysics toolbox (PTB-3) (Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) and EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002; Kleiner et al., 2007). Calibrations of eye fixations were conducted at the beginning of the experiment using a nine-point fixation procedure as implemented in the EyeLink API (see EyeLink Manual) and using Matlab software. Afterwards, calibrations were validated with the EyeLink software, and repeated when necessary until reaching an optimal calibration criterion. Drift correction was performed on each trial via central cross fixation.

In the *gender categorization task*, subjects were instructed to perform choice saccades towards female faces. In this task three images (viewpoint changes) for each PF individual (6 identities, 3 females) and their unfamiliar counterparts were presented. A trial began with a central fixation cross displayed between 800 and 1600 ms, followed by a 200 ms blank and subsequent presentation of the target/distractor pair presented for 600 ms. After a saccade was registered, the next trial was presented after a 1000 ms blank inter-trial interval. Stimuli subtended 14°x14° (average face height was 11°), and stimulus eccentricity was 8.6° of visual angle. With all possible combinations and equal number of presentations per identity and visual field, the total number of trials was 216; subjects took self-paced breaks after each block of 54 trials.

The two *familiarity categorization tasks* differed in terms of the number of identities depicted, but both required observers to perform choice saccades towards personally familiar identities presented with unfamiliar distractors. The low and high predictability variants involved presentation of 7 PF (3 females), or 3 PF (all male) identities, respectively, as well as an equal number of well-matched UF distractors. Note that, before beginning the respective experiments, participants were not aware of the number of identities actually presented. Presentation parameters were identical to those described for the gender categorization task (see above). The procedural parameters paralleled those used by Visconti di Oleggio Castello and Gobbini (2015), with exception of stimulus presentation duration (600 ms instead of 400 ms), as initial pilot testing revealed slightly longer presentation durations were necessary for acceptable performance levels. On each trial, a PF identity was paired with a same-gender, same-orientation distractor, and appeared with equal probability in either visual field. The total number of trials for the low predictability familiarity categorization task amounted to 150. To achieve a comparable number of trials in the high predictability variant for comparison with the low predictability variant, each unique stimulus × visual field combination was presented three times, leading to a total 162 trials over three blocks. These trials/blocks were doubled to further determine potential effects of repetition in the high predictability categorization task (see Analyses, Section 3). Subjects took self-paced breaks after each block of 50 or 54 trials (low or high predictability variant), respectively. Note that the order of the high and low predictability familiarity decision tasks was not randomized. Participants completed the low predictability version and the gender decision tasks (order counterbalanced) on the same day; two months later the high predictability variant was completed.

3. Analyses

3.1. Pre-processing

We applied the adaptive velocity based algorithm developed by Nyström and Holmqvist (2010) to find the onset of the first saccade (if any) within each trial. We discarded trials in which the onset of the first saccade was lower than 80 ms (Visconti di Oleggio Castello and Gobbini, 2015), as these were considered anticipatory saccades.

3.2. Statistical analyses

As mentioned above, across all experiments we considered only data from subjects whose performance exceeded chance level. For the gender categorization task this led to the exclusion of one departmental member ($n_{lab\_members} = 7/8$); all student participants performed above chance and were considered ($n_{students\_gender} = 36/36$). For the familiarity categorization tasks, only data from subjects who performed reliably across *both low and high predictability variants* were considered ($n_{lab\_members} = 4/8$, $n_{students\_familiarity} = 14/27$). Analyses performed to determine the effect of stimulus repetition were conducted on data from subjects who performed above chance across all blocks of the *high predictability* familiarity categorization task ($n_{lab\_members} = 7/8$, $n_{students\_familiarity} = 20/27$). Analyses of accuracy and mean SRTs were performed in R (version 3.2.4; R Core Team, 2013) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) and the lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015) to obtain p-values of the fixed predictors of the fitted models. Note that given our research question we were not interested in between-group differences, but rather those related to stimulus predictability, i.e. those observed *within* groups.

3.2.1. Gender categorization

Accuracy, mean and minSRTs are reported descriptively for the gender categorization task for all participants ($n_{lab\_members,}$ $n_{students\_gender}$), as this task served only as a baseline to demonstrate subjects' ability to perform the SRT task.

3.2.2. Personal familiarity categorization: low vs. High stimulus predictability

Accuracy. To investigate the effect of stimulus predictability on subjects' accuracy, we performed generalized linear mixed models with a binomial family (Jaeger, 2008) for the data obtained in the experiments characterized by lower (7 identities), or higher stimulus predictability (3 identities), respectively. This was done separately per group tested given the unequal sample sizes available. In this model, the main predictor is the variable 'predictability' (*low* and *high* for larger and smaller number of identities presented) and the variable participant is a random factor. We performed the Log-likelihood Ratio Test to compare the null and full model and assess the significance of the predictor.

Mean SRTs. To investigate the effect of stimulus predictability on SRTs, we performed a linear mixed model for the data obtained in the experiments characterized by lower (7 identities), or higher stimulus predictability (3 identities), respectively, considering only the correct trials. Note that, due to the different number of trials across these experiments, we considered all trials from the low predictability (1 5 0), and those of first 3 blocks for the high predictability (1 6 2) familiarity categorization tasks. In this model, the main predictor is the variable 'predictability' and the variable participant is a random factor. As for accuracy scores, we performed the Log-likelihood Ratio Test to compare the null and full model and assess the significance of the predictor.

Minimum SRTs. We estimated minSRTs in two different ways, here (as for mean SRTs) also only considering the first three blocks of trials. In order to facilitate comparison across studies, we opted for the same procedures as reported by Besson et al. (2017) and Visconti di Oleggio Castello and Gobbini (2015). First, across subjects' trials (i.e., *group minSRTs*) we performed a chi-square test using 10 ms time bins across trials. We considered the first bin where the number of correct trials outperforms statistically the number of incorrect trials ($p < .05$), followed by at least three significant consecutive bins (Besson et al., 2017). Second, we determined *individuals' minSRTs*. To this end, we (i) considered only RTs of participants who performed above chance level, and (ii) opted for 40 ms time bins using the Fisher's exact test ($p < .05$). Using this procedure, some participants' individual minSRT could not be computed (due to proximal in/correct trial distributions).

Table 1
Accuracy (in %) and minimum saccadic reaction times (in ms) obtained across experiments for all subjects tested, whose performance was above chance level.

Gender categorization		
	Department members ($n_{lab\_members=7}$)[1]	Students ($n_{students\_gender=36}$)
Accuracy	81 [71, 90]	84 [81, 86]
minSRT	200 ms	140 ms
Mean SRT	276	252
Median SRT	262	236
CI	[272, 281]	[251, 254]
Personal familiarity categorization		
	Department members ($n_{lab\_members=4}$)	Students ($n_{students\_familiarity=14}$)
Low predictability		
Accuracy	74 [64, 81]	68 [64, 72]
minSRT	360	260
Mean SRT	373	334
Median SRT	390	334
CI	[370, 388]	[331, 340]
High predictability		
Accuracy	82 [74, 89]	76 [71, 79]
minSRT	260	240
Mean SRT	345	345
Median SRT	353	339
CI	[343, 357]	[343, 350]

95% confidence intervals are provided in brackets. Individual observers' data are reported in Table S1 and Table S4.

[1] Considering only the four obsservers reported for personal familiarity categorizations yielded the following performance: Accuracy: 91% [84 94]; MinSRT: 200 ms; Mean SRT: 305 ms; Median SRT: 291 ms; CI [299, 310].

Finally, we assessed the effect of stimulus set size for *individual subjects* of n_3 using the Wilcoxon signed-rank test ($p < .05$). Note that the comparison of minSRTs as a function of stimulus set size was not conducted for n_1 due to the insufficient statistical power, as only four participants' data were considered.

3.2.3. Effect of stimulus repetition under conditions of high stimulus predictability

To determine the effect of stimulus repetition, we used the above described procedure for all behavioral measures, however, taking into account *all* correct trials (from *all 6 blocks* of the high predictability familiarity categorization experiment; cf. above). We compared RT associated with the first presentation of a stimulus with each subsequent presentation.

4. Results

Table 1 summarizes the results obtained for all subjects for gender categorization ($n_{lab\_members}$, $n_{students\_gender}$), and personally familiar face recognition ($n_{lab\_members}$, $n_{students\_familiarity}$) under low and high predictability conditions, respectively. Fig. 2 shows individuals' minSRTs plotted against accuracy scores across all experiments. For Individual subjects' data see Supplemental Tables.

4.1. Gender categorization

Both departmental members and student participants could reliably perform the gender categorization task, achieving 81% and 84% on average, and exhibiting minimum SRTs of 200 ms and 140 ms, respectively (see Fig. 3).

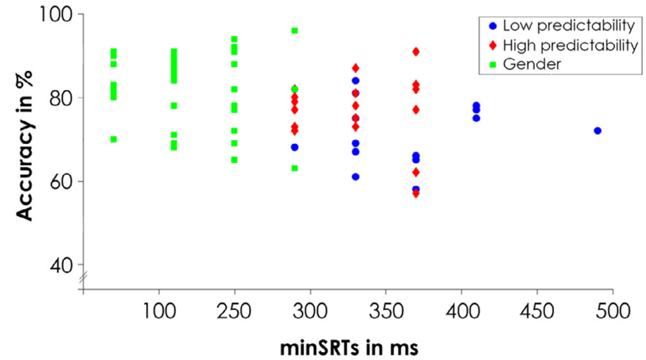

Fig. 2. Individual subjects' minSRTs plotted against performance accuracy. Note that for 3 student subjects minSRTs could not be computed (cf. Table S4).

4.2. Personal familiarity categorization: lower vs. Higher stimulus predictability

Accuracy. The groups' SRT distributions for the low and high predictability familiarity categorization tasks are illustrated in Fig. 4. For departmental members' accuracy scores, the fitted model (considering the total of 1248 trials) revealed a significant main effect of stimulus set size (i.e., number of identities depicted; $X^2(1) = 13.73$, $p = .0002$); performance given 7 target identities was significantly lower than when 3 target identities were presented (74% vs. 82%; z = 3.69, $p = .0002$). For students' accuracy scores, the results from the fitted model (based on a total of 4361 trials) also showed a significant main effect of stimulus set size $X^2(1) = 33.53$, $p < .0001$); again, performance for the lower as compared to higher predictability task variant was significantly lower (68% vs. 76%; z = 5.78, $p < .0001$). For parameter estimates of the fixed effects for the generalized linear mixed model with the binomial family for each group tested see Table S2.

Mean SRTs. For departmental members' SRTs the fitted model (considering the total of 967 trials) revealed a significant main effect of stimulus set size (X^2 (1) = 29.70, $p < .0001$). Mean (373 vs 345; t = −5.49, $p < .0001$). For students' SRTs the results from the fitted model (based on a total of 3142 trials) showed a significant main effect of stimulus set size ($X^2(1) = 14.67$, $p = .0001$). Contrary to the departmental members, the mean of RTs in Experiment 1 was significantly faster than in Experiment 2 (334 vs 345; t = 3.83, $p = .0001$). For parameter estimates of the fixed effects for the Linear Mixed Model for each group tested see Table S3.

MinSRTs. The Wilcoxon signed-rank test revealed no main effect of stimulus set size on the *individual* minSRTS (see Table S4) for student participants in $n_{students\_familiarity}$ ($p = .11$).

4.3. Effect of stimulus repetition under conditions of high stimulus predictability

Accuracy. The fitted model (considering the total of 1944 trials) revealed no significant effect of repetition on subjects accuracy scores for n_1 ($X^2(5) = 2.14$, $p = .83$). However a significant main effect of repetition was shown for n_2 ($X^2(5) = 16.22$, $p < .05$; based on 6480 trials in total). Performance in the first image presentation (73%) was significantly lower than for the fourth image presentation (z = 2.61, $p < .05$) and the fifth image presentation (z = 3.38, $p < .001$) with a performance of 78% and 79%, respectively (see Table 2). For parameter estimates of the fixed effects for the generalized linear mixed model with the binomial family for each group tested see Table S5.

Mean SRTs. The fitted model revealed a significant effect of repetition on mean RTs for n_1 ($X^2(5) = 32.18$, $p < .0001$) and n_2 ($X^2(5) = 126$, $p < .0001$) (based on a total 1411 and 4834 trials, respectively). Both groups were slower in the first image presentation compared to the five other image presentations (see Table 2). For

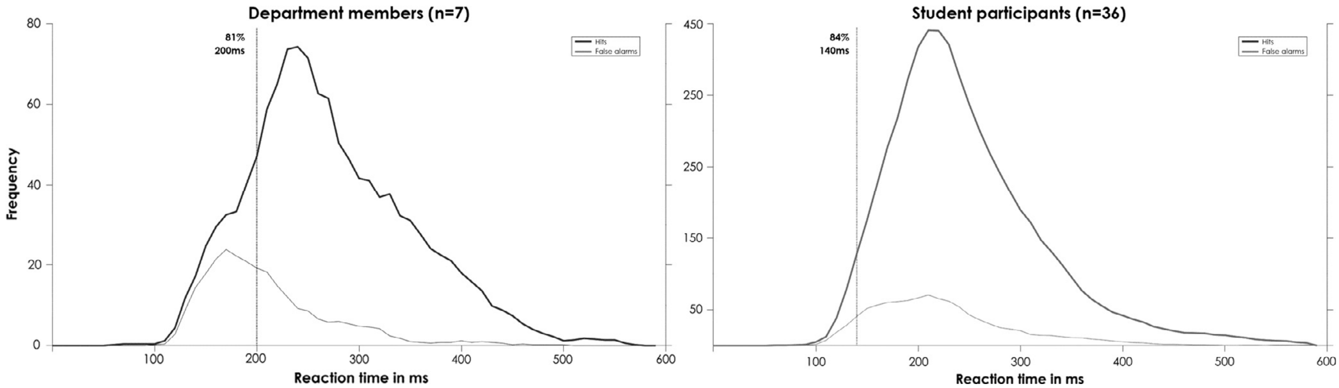

Fig. 3. Distributions of participants' SRTs expressed during the gender categorization task. Hits and false alarms per time bin are indicated as thick and thin lines, respectively. Vertical lines indicate each group's min SRTs; along with average accuracy.

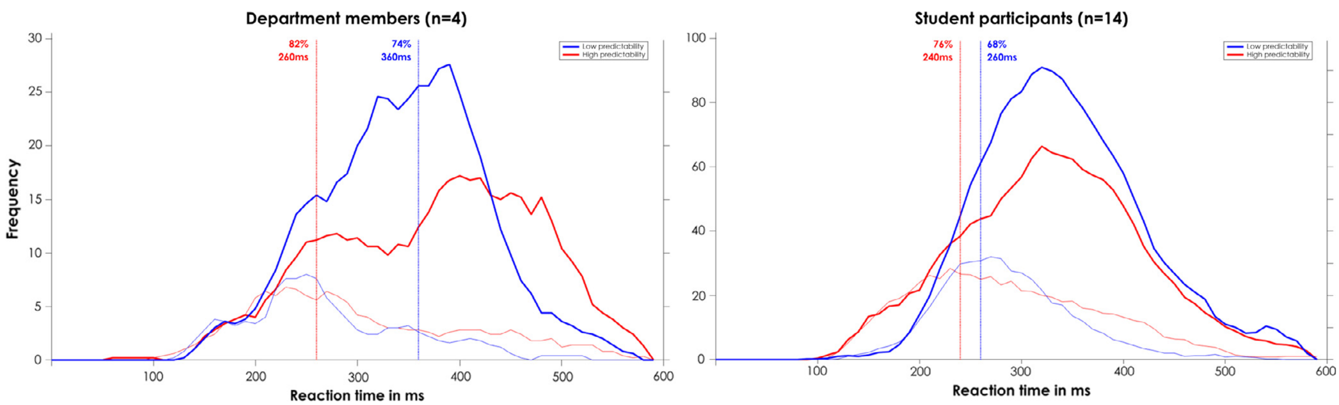

Fig. 4. Distributions of participants' SRTs expressed during the personal familiarity categorization task with low (blue) and high (red) predictability. Hits and false alarms per time bin are indicated as thick and thin lines, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

parameter estimates of the fixed effects for the Linear Mixed Model for each group tested see Table S6.

5. Discussion

Human vision is a sensory system that is extensively investigated to advance our understanding of cognition and brain functioning in general. Similar to other sensory modalities, visual processing enables extremely robust categorization of external stimuli. Previous work has demonstrated that the primate brain can for instance visually categorize faces and animals in a highly proficient and rapid manner (e.g. Crouzet et al., 2010; Fabre-Thorpe, 2011; Kirchner & Thorpe, 2006; Thorpe et al., 1996; Rousselet et al., 2003). Visual categorization paradigms, which are considered to involve activation of representations that impact observers' responses via maximal "presetting" of the visual system for the task at hand , have been deployed to constrain theories of visual processing (for a review see e.g., Fabre-Thorpe, 2011). For the most, previous studies have focused on how visual categorization speed stimulus varies as a function of response effector (e.g., Bacon-Macé et al., 2007), stimulus category (e.g., Crouzet et al., 2010), processing level or task- dependent information and diagnosticity (e.g., Schyns, 1998).

In the present study, we sought to systematically investigate the extent to which visual categorization speed of complex visual stimuli is affected by processing level and observers' expectations, which can vary depending on stimulus set size (cf., Ramon, 2018; Ramon et al., 2018; Ramon & Rossion, 2010; Ramon, Busigny, Gosselin & Rossion, 2017; Ruffieux et al., 2017). To address this, we measured the speed of gender, and familiar face categorization in a SRT paradigm (cf. Ramon et al., 2018). Moreover, to address previously reported incompatible findings regarding the speed of familiarity categorization (cf. Visconti

Table 2
Accuracy, mean and min SRTs for choice saccades towards personally familiar faces as a function of stimulus repetition during familiarity decisions under high predictability.

		Department members ($n_{lab\_members}$)				Students ($n_{students\_familiarity}$)		
Block	Accuracy in %	Mean SRTs	CI	minSRTs	Accuracy in %	Mean SRTs	CI	minSRTs
1st	77	331	[277; 386]	250	73	350	[325; 374]	280
2nd	78	307	[239; 375]	300	74	330	[298; 361]	260
3rd	75	306	[238; 375]	340	76	326	[294; 357]	240
4th	76	293	[225; 361]	290	78	312	[280; 343]	220
5th	74	310	[241; 379]	260	79	318	[286; 350]	220
6th	74	306	[237; 374]	290	74	319	[287; 351]	240

Note that per block each stimulus was shown three times.

di Oleggio Castello & Gobbini, 2015; vs. Besson et al., 2017; Mathey et al., 2012; Ramon et al., 2018), we varied the number of personally familiar face target identities to be detected (3 vs. 7, along with an equal number of distractors). We reasoned that, beyond reconciling these apparent inconsistencies in the face literature, our anticipated findings of processing level and stimulus set size dependent categorization would be relevant for both visual categorization specifically, and cognition in general.

In line with previous findings (Ramon et al., 2018; cf. also Besson et al., 2017), visual categorization proficiency varied with processing level. Gender categorizations were performed more accurately and rapidly than familiarity categorizations. Interestingly, gender categorization latencies reported here and elsewhere (Ramon et al., 2018) were comparable to those expressed when observers have to saccade towards only a single target identity (140 ms; Mathey et al., 2012). This is the first line of evidence suggesting that it is not just the processing level, task, or stimulus familiarity *per se* that determines visual categorization speed.

Additionally, extending previous work (e.g., Miller, 1956; Palmer, 1994; Busey & Palmer, 2008; Kent & Lamberts, 2005; Lacouture & Marley, 1995; Karpiuk et al., 1997), stimulus set size affected performance expressed in the context of the same processing level, i.e. during familiarity categorizations. Specifically, categorization was faster and/or more accurate for smaller, as compared to larger stimulus sets. Saccades could be directed towards more predictable targets (three familiar identities) within ~250 ms, although lab members performed more accurately than students. Increasing the number of targets did not affect students' mSRTs, but led to a decrease in their accuracy (76% vs. 68%). Lab members, on the other hand, responded slower (360 ms), and their performance was less accurate (82% vs. 74%).

The results reported here demonstrate that both processing level, and decisional space, or stimulus set size affect categorization proficiency. Our findings demonstrate that it is not "detection of personal familiarity" (Visconti di Oleggio Castello & Gobbini, 2015) per se that expedites categorization, but the experimental circumstances that enable the measured latencies. Note that individuals can be differently affected by such procedural choices, with some observers' performance varying more in terms of response latencies, or accuracy, respectively. Therefore, in order to determine the boundaries of visual categorization speed, simply establishing mSRTs is insufficient.[2] At least in the context of categorization of complex visual stimuli, processing should not be considered as fixed, but rather flexibly modulated through a combination of contextual and structural aspects, related to both procedural choices (task, processing level, stimulus set) as well as the stimulus category (e.g., Lipp et al., 2015; Craig et al., 2012; Schyns, 1998). Disregarding such aspects can result in seemingly inconsistent findings (cf. Lipp et al., 2015), spur scientific debates, and promote erroneous conclusions. The challenge is therefore to carefully consider the conditions under which categorization proficiency is observed, and adopt the appropriate terminology to communicate the findings made (cf. Ramon, 2018; Ramon et al., 2018; Ramon, in press; Ramon et al., in press).

6. Conclusion

Our findings demonstrate the powerful effect that both processing level and decisional space, as varied through stimulus set size, can exert on visual categorization speed. Based on the present and previous findings we emphasize the importance of adopting appropriate terminology that respects the level of processing or task type (Besson et al., 2017), stimulus repetition (Lewis & Ellis, 2000; Ramon et al., 2011), and decisional space within which visual categorizations are performed (Ramon et al., 2018; Ruffieux et al., 2017), which can account for contradictory findings and conclusions (Ramon, 2018; Lipp et al., 2015; Ramon et al., 2018). To make this point, we demonstrate that categorization speed modulation, which was previously attributed to real-life familiarity (Visconti di Oleggio Castello & Gobbini, 2015), can in fact be more parsimoniously accounted for in terms of observers' expectations. We advocate for careful consideration of procedural aspects when categorization proficiency serves to constrain theories of visual processing, and cognition more generally.

[2] Consider for instance the very rapid mSRTs reported previously on the group level for 3-target familiarity categorizations (180ms; Visconti di Oleggio Castello & Gobbini, 2015). The same task could be achieved in as little as 150ms by an individual subject, who performed slightly, but significantly above chance (Ramon et al., 2018).

Acknowledgements

We thank our previous Bachelor students Nour Al-Khodairy, Morgane Bulliard, Aurelie Christinaz, and Lysiane Constantin for their invaluable help during data collection.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.01.019.

References

Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1013–1026.

Bacon-Macé, N., Macé, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45*(11), 1459–1469.

Barragan-Jason, G., Cauchoix, M., & Barbeau, E. J. (2015). The neural speed of familiar face recognition. *Neuropsychologia, 75*, 390–401.

Barragan-Jason, G., Lachat, F., & Barbeau, E. J. (2012). How fast is famous face recognition? *Frontiers in Psychology, 3*, 454. https://doi.org/10.3389/fpsyg.2012.00454.

Barton, J. J. S., & Corrow, S. L. (2016). Recognizing and identifying people: A neuropsychological review. *Cortex. 75*(Feb), 132–150. https://doi.org/10.1016/j.cortex.2015.11.023.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. < http://CRAN.R-project.org/package=lme4 >.

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

Busey, T., & Palmer, J. (2008). Set-size effects for identification versus localization depend on the visual search task. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 790–810. https://doi.org/10.1037/0096-1523.34.4.790.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS One, 3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci. 34*(13), 4548–4557. https://doi.org/10.1523/JNEUROSCI.4339-13.2014.

Caharel, S., Ramon, M., & Rossion, B. (2014). Face familiarity decisions take 200 msec in the human brain: Electrophysiological evidence from a go/no-go speeded task. *Journal of Cognitive Neuroscience, 26*(1), 81–95. https://doi.org/10.1162/jocn_a_00451.

R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: < http://www.R-project.org/ >.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617. https://doi.org/10.3758/BF03195489.

Craig, B. M., Mallan, K. M., & Lipp, O. V. (2012). The effect of poser race on the happy categorization advantage depends on stimulus type, set size, and presentation duration. *Emotion. 6*, 1303–1314. https://doi.org/10.1037/a0028622.

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2009). The bear before the forest, but the city before the cars: Revealing early object/ background processing. *Journal of Vision, 9*, 954.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades towards faces: Face detection in just 100 ms. *Journal of Vision, 10*(16), 1–17.

Caudek, C. (2013). The fidelity of visual memory for faces and non-face objects. *Acta Psychologica (Amsterdam), 143*(1), 40–51. https://doi.org/10.1016/j.actpsy.2012.12.002.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*(4), 576–585.

Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by FMRI. *Cerebral Cortex, 17*(9), 2123–2133.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2*, 243. https://doi.org/10.3389/fpsyg.2011.00243.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport, 9*, 303–308.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261–2271.

Hall, M. G., Mattingley, J. B., & Dux, P. E. (2018). Electrophysiological correlates of incidentally learned expectations in human vision. *Journal of Neurophysiology, 119*(4), 1461–1470. https://doi.org/10.1152/jn.00733.2017.

Hasbroucq, T., Mouret, I., Seal, J., & Akamatsu, M. (1995). Finger pairings in two-choice reaction time taskscolon: does the between-hands advantage reflect response preparation? *Journal of Motor Behavior, 27*(3), 251–262.

Huang, W., Wu, X., Hu, L., Wang, L., Ding, Y., & Qu, Z. (2017). Revisiting the earliest electrophysiological correlate of familiar face recognition. *International Journal of Psychophysiology, 120*, 42–53.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.

Kadel, H., Feldmann-Wüstefeld, T., & Schubö, A. (2017). Selection history alters attentional filter settings persistently and beyond top-down control. *Psychophysiology, 54*(5), 736–754. https://doi.org/10.1111/psyp.12830.

Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology Learning Memory and Cognition. 31*(2), 289–305.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*, 1762–1776.

Kim, N. Y., & McCarthy, G. (2016). Task influences pattern discriminability for faces and bodies in ventral occipitotemporal cortex. *Society Neuroscience, 11*(6), 627–636. https://doi.org/10.1080/17470919.2015.1131194.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement), https://doi.org/10.1068/v070821.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2015). Package 'lmerTest'. R package version, 2(0).

Lacouture, Y., & Marley, A. A. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & Psychophysics, 66*(7), 1206–1226.

Lewis, M. B., & Ellis, H. D. (2000). The effects of massive repetition on speeded recognition of faces. *Quarterly Journal of Experimental Psychology Section A, 53*, 1117–1142.

Lipp, O. V., Karnadewi, F., Craig, B. M., & Cronin, S. L. (2015). Stimulus set size modulates the sex-emotion interaction in face categorization. *Attention, Perception, & Psychophysics, 77*(4), 1285–1294. https://doi.org/10.3758/s13414-015-0849-x.

Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience, 21*, 2007–2018.

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General, 144*(3), 551–569. https://doi.org/10.1037/a0039184.

Manahova, M. E., Mostert, P., Kok, P., Schoffelen, J. M., & de Lange, F. P. (2018). Stimulus familiarity and expectation jointly modulate neural activity in the visual ventral stream. *Journal of Cognitive Neuroscience, 30*(9), 1366–1377. https://doi.org/10.1162/jocn_a_01281.

Mathey, M. A., Besson, G., Barragan-Jason, G., Garderes, P., Barbeau, E. J., & Thorpe, S. J. (2012). Sarkozy: Left or right? How early can we choose? *Perception ECVP Abstract Supplement, 41*, 166.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods, 42*(1), 188–204 doi:0.3758/BRM.42.1.188.

O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B., Jr, & Bar, M. (2017). Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Consciousness and Cognition, 47*, 63–74. https://doi.org/10.1016/j.concog.2016.05.003.

Olk, B., & Garay-Vado, A. M. (2011). Attention to faces: Effects of face inversion. *Vision Research, 51*(14), 1659–1666. https://doi.org/10.1016/j.visres.2011.05.007.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Ramon, M. (2018). The power of how – lessons learned from neuropschology and face processing. *Cognitive Neuropsychology*.

Ramon, M., Caharel, S., & Rossion, B. (2011). The speed of recognition of personally familiar faces. *Perception, 40*(4), 437–449.

Ramon, M., & Gobbini, M. I. (2018). Familiarity matters: A review of prioritized processing of personally familiar faces. *Visual Cognition, 26*, 179–195.

Ramon, M., Bobak, A. K., & White, D. (2018). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology*.

Ramon, M., Sokhn, N., Lao, J., & Caldara, R. (2018). Decisional space determines saccadic reaction times in healthy observers and acquired prosopagnosia. *Cognitive Neuropsychology*.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*, 629–630.

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*, 440–456.

Ruffieux, N., Ramon, M., Lao, J., Colombo, F., Stacchi, L., Borruat, F. X., ... Caldara, R. (2017). Residual perception of biological motion in cortical blindness. *Neuropsychologia, 93*, 301–311.

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition, 67*(1–2), 147–179.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457–482.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra- rapid visual categorisation of natural and artifactual objects. *Perception, 30*, 655–668.

Visconti di Oleggio, M., Castello, & Gobbini, M. I. (2015). Familiar Face Detection in 180 ms. *PLoS ONE, 10*(8), e0136548. https://doi.org/10.1371/journal.pone.0136548.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.67.