Abstract

The question of whether age and facial identity are processed through a shared or parallel-route has scarcely been examined, despite being of theoretical relevance for face processing models. For the first time, the Garner speeded classification paradigm was applied to assess the independence of age and facial identity processing. Across three experiments, participants made either age or facial identity judgements while both dimensions vary (Filtering) or only one dimension varies while the other remains constant (Baseline). Garner interferences, represented by slower response times for the Filtering condition compared to the Baseline condition, were recorded for both Experiments 1 (familiar, cropped, single-image stimuli) and 2 (unfamiliar, cropped, single-image stimuli). A weaker Garner interference was recorded for Experiment 3 (familiar, naturalistic, multi-image stimuli). Garner interference for the first two experiments is indicative of the shared route hypothesis for identity and age perception. However, findings from Experiment 3 suggest that these effects are weaker for naturalistic images and the implications of this are discussed.

Keywords: Garner speeded classification paradigm; facial identity; facial age; age perception; face perception

Introduction

Traditional theoretical face processing models (Bruce & Young, 1986) considered judgements of age to be one of several *visually derived semantic codes* defined as information which can be inferred from visual cues with relative accuracy. These visually derived semantic codes, which also include sex and attribute judgements, were considered independent by-products of the operation of face recognition but not essential components for face recognition. However, the theory that these processes are independent has since been challenged, suggesting for example that sex judgements and facial identity are processed through a single-route (Ganel & Goshen-Gottstein, 2002). Age perception, within Bruce and Young's (1986) theoretical framework, was not considered any further and subsequent theoretical models of face processing do not include age (Burton et al., 1999; Haxby et al., 2000; Valentine, 1991). Yet, the face undergoes progressive transformations across the lifespan altering the structure of the face, facial features, and skin texture and tone. However, some idiosyncratic physical properties are retained, and these can be used in the recognition of familiar faces from images taken decades apart (Mileva et al., 2020). This suggests that age may in fact be an important component of face processing, and it is not yet established whether or to what extent face recognition and age perception share a single processing route.

The Interaction Activation and Competition (IAC) model (Burton et al., 1999) proposes that the architecture of face processing is structured around pools of units which connect bidirectionally to units in other pools. This structure comprises a Face Recognition Unit (FRU), Person Identity Nodes (PIN), Semantic Information Unit (SIU), and Word Recognition Unit (WRU). There is one FRU for each face and this is view-independent, while one PIN unit contains information on a person, rather than the face.

At PIN, the model proposes that other known aspects of the person converge, such as, the name (WRU) and occupation (SIU). The IAC does not account for the perception of age within this architectural structure. It can be argued that age would be stored in the Semantic Information Unit such that known information about the age of an identity can be drawn once the PIN is activated. However, unlike the name, occupation, and other semantic information, age is a fundamental visual feature which is relatively unstable and transforms slowly over time. Although age could be comparable to other visually derived facial information, specifically sex and emotional expression, it is also unique. Sex is typically stable over time, and an observer can gain exposure to a range of emotional expressions within a single meeting. In contrast, different ages of a person are not typically viewed during the learning phase of a face, such as in the example of a single encounter. Therefore, in comparison to exposure to emotional expression, exposure to different ages of the same identity is more infrequent in everyday interactions.

An alternative theory, as suggested by the Bruce and Young model (1986), is that age perception, being a visually derived semantic code with no functional purpose is just a by-product of facial recognition. From a functional perspective, forming an age-independent representation of a face would be beneficial since it would allow for a robust facial recognition processing system which is not impaired by age-related facial changes over time. Computer recognition models have been shown to recognise familiar faces from images depicting the face decades apart and stable idiosyncratic features might be key to recognising faces despite these aging transformations (Mileva et al., 2020). However, in human perception it is not clear to what extent independent representations of facial identity, such as idiosyncratic features, or features associated with fundamental aspects of facial identity recognition, contribute to recognition of identities across ages. There is some overlap in low-level perceptual features which have been identified as important for both facial identity recognition and age perception, such as pigmentation (i.e., surface reflection, luminance, and colour contrast), and alterations to these features due to aging can also change representation of identity (Galper, 1970; Johnston & Edmonds, 2009; Porcheron et al., 2013; Russell et al., 2006;). This suggests that aspects of the face which are important for face recognition are also changeable, at least to some degree, during the aging process. However, the unnatural image distortions created by the experimental manipulations implemented in such studies (e.g., using photographic negatives) make generating inferences about overlap between key features used in facial identity recognition and age perception challenging. Nonetheless, a bottom-up account of crosstalk has been reported between other aspects of face perception, specifically for example perception of gender and age because of shared features underlying both these mechanisms (Fitousi, 2021). Additionally, there may also be top-down influences whereby representations of age and identity overlap at the level of prior knowledge, stereotypes, and social expectations, or a dynamic interplay between the two accounts as described in Freeman and Ambady's (2011) model of person construal. Thus, for example, in the context of age and identity, knowledge of the release date of an actor's movie can enhance processing of their current age.

Therefore, it is not inconceivable that the processing of age shares some underlying mechanism with facial identity processing. However, Haxby et al's (2000) neural model of face perception proposes that the brain networks for changeable aspects of the face are independent to invariant aspects of the face. In this model, changeable aspects of face perception refer to the perception of movement and information that facilitates social communication, such as gaze direction, emotional expressions, and facial speech. In contrast, invariant aspects are associated with stable dimensions, such as facial identity, gender, and race. The Posterior Superior Temporal Sulcus is responsible for processing changeable information, while the Fusiform Gyrus is used for the invariant face information (Haxby et al., 2000; Hoffman & Haxby, 2000; Kanwisher et al., 1997), although the degree of separation between the functional roles of the different brain regions is not clear (Haxby et al., 2000).

Age, however, is not directly considered within Haxby et al's (2000) neural system and whether age should be classified as a changeable or invariant aspect of the face is not specified. Although there is no dispute that age does involve significant changes to the face (Porcheron et al., 2013), in many ways it is unlike other changeable aspects considered in the model. Specifically, the transformations to the face due to aging occur over a longer period and, although age provides social context and impacts on social interactions, it does not primarily function to convey social information in the same direct and fleeting way as emotional expression, facial speech movements, and gaze direction.

Parallel or Shared Route? Insights from Empirical Studies

Taken together, theoretical models of face processing often do not directly account for the perception of age. If age information is to be considered within the same category as other changeable non-identity face information, theoretical models would suggest that age and facial identity operate independently. A small set of cognitive and behavioural studies provide evidence for the independence hypothesis. Examination of the dissociation of ability between different face perception tasks in those with impaired facial recognition ability suggests that ability to extract other non-identity facial information remains intact (e.g., Chatterjee & Nakayama, 2012; Nunn et al., 2001).

Specifically, for example, individuals with developmental prosopagnosia do not differ from the control group in their ability to order computer generated unfamiliar faces from young to old (Chatterjee & Nakayama, 2012). These studies provide some support for the independent route. However, the ability to estimate age using different tasks, natural images, and varying difficulty levels was not assessed (Chatterjee & Nakayama, 2012). Furthermore, the nature and severity of face processing impairment may affect the range of cognitive, perceptual, and face-related processes impacted (Bate et al., 2019; Behrmann et al., 2005; Le Grand et al., 2006; Rosenthal & Avidan, 2018). This makes drawing conclusions with respect to the (in)dependence of age and facial identity from research with this population challenging.

In contrast, other behavioural studies challenge the notion that age perception and facial identity operate parallel-routes. In a repetition priming study, the ages of primed familiar and famous faces were classified faster and more accurately than non-primed faces (Dagovitch & Ganel, 2010). Priming was achieved by asking participants to provide judgements of the face images (attractiveness, trustworthiness, pleasantness ratings) during the priming phase which was then followed by rating the age of these faces (Dagovitch & Ganel, 2010). However, there is some debate concerning whether patterns recorded in these studies reflect mere exposure effects or true repetition priming (for review see Butler, 2004). This caveat limits the conclusions that could be drawn about (in)dependence of familiar face processing and age perception. Furthermore, it provides little indication as to the interaction of age perception with *unfamiliar* face processing. An applied study examining unfamiliar face matching and age verification on accuracy when checking ID-cards (Robertson & Burton, 2021) suggests that there may be an interaction between age perception and unfamiliar face processing. A striking interference between the two tasks was recorded, such that the task of verifying age by a neurotypical population resulted in an increase in face matching error by a margin of 46%, and subsequent experiments excluded an increase in cognitive load as the reason (Robertson & Burton, 2021). Although this study did not directly assess independence in processing, the interference in performance suggests that there may be a degree of interaction.

Studies vary in their use of familiar and unfamiliar faces, however there is now a substantial body of evidence demonstrating that familiar and unfamiliar faces are processed differently (for review see Johnston & Edmonds, 2009), and findings from familiar and unfamiliar faces should not be conflated (Burton, 2013). In the context of age processing, estimating the age of familiar faces may involve some degree of retrieval of semantic information related to any prior knowledge the observer holds about the identity (Bruce & Young, 1986). Therefore, perception of age of familiar faces across the lifespan may be supported by an additional top-down component not available for unfamiliar faces. In fact, when comparing familiar and unfamiliar faces on an age decision task, familiar faces recorded a significant accuracy advantage (Bruyer et al, 1991; Bruyer et al., 2007). Although this effect was reliant on perceptual difficulty and present only when the faces were presented for brief exposures, it does suggest that familiar and unfamiliar faces may interact differently with age processing.

In sum, although previous face perception models account for processing of non-identity aspects of face perception (e.g., sex, race, emotional expression) (Bruce & Young, 1986; Burton et al., 1999; Haxby et al., 2000; Valentine, 1991), age and how it interacts with facial identity processing is not accounted for in any direct or complete way. Considering age to be like other non-identity aspects of face processing is also problematic. This is because of the slow progressive transformation which is a unique characteristic to age not shared with any of the other aspects of face processing, such as race which is stable over time or emotional expression which changes frequently over just a few seconds. Many insights are gleamed indirectly through studies examining other aspects of face processing (Chatterjee & Nakayama, 2012; Nunn et al., 2001; Robertson & Burton, 2021), and few studies have made direct attempts to understand whether facial identity processing and age perception are processed through parallel or shared routes (Dagovitch & Ganel, 2010).

The findings from these studies are inconsistent and the question of whether identity and age processing operate a shared or parallel-route remains unresolved leaving a gap in our understanding of age processing within existing face processing frameworks. Therefore, over a series of three experiments[1], the present study will systematically examine the question of independence of facial identity and age processing for both familiar and unfamiliar faces using the Garner speeded classification paradigm. This paradigm has been used previously to assess independence of different dimensions of face processing (see Algom & Fitousi, 2016), but has not yet to be applied to examine facial identity and age processing.

The Garner Speeded Classification Paradigm

The Garner speeded classification paradigm, is a well-established approach for examining the independence of wide range of perceptual dimensions, including shape perception, numerical cognition, auditory perception, and face perception (see for example, Algom & Fitousi, 2016; Atkinson et al., 2005; Ganel & Goshen-Gottstein, 2002; Schweinberger & Soukup, 1998; Wang et al., 2013). This paradigm is based on a

[1] *Note.* Two additional experiments originally included in the pre-registered report have been omitted from the study. The two experiments proposed a manipulation of Relative Baseline Discriminability by morphing the face stimuli of one dimension to create mismatched discriminability. Data was not collected due to challenges with achieving suitable level of error during piloting of stimuli.

theoretical concept of selective attention that is, the ability to focus on relevant stimuli while suppressing responses to irrelevant or conflicting information. Therefore, in this paradigm, participants selectively attend to one aspect of the stimulus (i.e., the task-relevant dimension) while ignoring another aspect (i.e., the task-irrelevant dimension) (see Algom & Fitousi, 2016; Garner, 1976; Fitousi, 2023). The theoretical concept is similar to that of the Stroop effect, whereby a failure to selectively attend to a single stimulus dimension suggests that the stimulus is composed of dimensions which interact in processing, i.e., *integral* dimensions. Thus, integral dimensions are processed as a unified whole, and examples of integral dimensions include colour ink and colour word (see Eidels et al., 2010), hue and saturation (Garner, 1974), loudness and timbre (Melara & Marks, 1990). In contrast, successful selective attention suggests that the dimensions of the stimulus do not interact in processing, i.e., *separable* dimensions (for a detailed description see, Algom & Fitousi, 2016). Thus, separable dimensions can be deconstructed and processed independently, and examples of separable dimensions include, colour and shape (Garner, 1974), size and brightness (Garner, 1977), and colour and texture (Cant et al., 2008). Converging evidence from other methodologies provide further support for the distinction between integral and separable dimensions (see Algom & Fitousi, 2016). However, a detailed account of the mechanisms underlying Garner Interference and its relationship to other methodologies is beyond the scope of this paper but comprehensive reviews and critical evaluations can be found in the works of Algom & Fitousi (2016), Fitousi (2023), and Melara (1993).

This paradigm usually comprises three blocks of trials. In the Baseline condition, the task-irrelevant dimension is held constant while the task-relevant dimension varies. In the Filtering condition, the task-irrelevant dimension will vary randomly. In the correlated condition, the task-irrelevant dimension will vary consistently across all trials with the relevant dimension. Taking emotional expression and facial identity as an example. If the task here is to classify facial identity while ignoring emotional expression, the task-relevant dimension is the facial identity of the stimuli, and the task-irrelevant dimension is emotional expression. In the Baseline condition, the identity of faces (i.e., task relevant) will differ, but facial expression (i.e., task irrelevant) will be held constant across all trials in this block (e.g., Persons A and B always express anger). In the Filtering condition, the task-irrelevant facial expression dimension will vary *randomly* (e.g., Persons A and B randomly express anger or disgust). In the correlated condition, changes to the facial expression will vary *consistently* with changes to facial identity (e.g., Person A always expresses anger and Person B always expresses disgust). The dimensions can be replaced with other aspects of the face, including facial identity and facial age, which are relevant to this paper. A graphical illustration of the Baseline, Filtering, and Correlation blocks using facial identity and age as dimensions is presented in Figure 1.

The difference in response times across these three conditions reveals the success of selective attention to a single aspect of the stimulus. A measure of *Garner interference* is calculated by computing the difference in mean response times between the Baseline and Filtering conditions. Slower response times in the Filtering condition would indicate that the task-irrelevant dimension influenced the ability to selectively attend to the task-relevant dimension. An analogous calculation is also often performed to calculate the difference in mean response times between the correlated condition and the baseline condition. If response times for the correlated condition are faster than the baseline condition, this is called *Redundancy Gain.* The two dimensions are characterised as *separable* if speed and accuracy of performance is comparable across all three conditions. In contrast, if the results reveal both Garner interference and

Redundancy gain, then the dimensions are considered *integral* (see Algom & Fitousi, 2016).

Another possible outcome is asymmetrical interference. This term describes a scenario where one dimension is found to influence a second dimension, but the reverse is not found. For example, facial identity has been found to interfere with emotion expression classification, but emotion expression does not interfere with facial identity recognition (Schweinberger & Soukup, 1998). Although Garner suggested that this was likely the result of a hierarchy in processing, rather asymmetry may be determined by baseline discriminability (see Algom & Fitousi, 2016). *Discriminability* is defined as the difference separating two stimuli along a single dimension (Melara & Mounts, 1993). Baseline discriminability is matched across both dimensions when the recorded speed and accuracy is comparable for the two dimensions in the baseline condition. The effect of matched and mismatched baseline discriminability between the two dimensions has been examined through a series of studies using a range of perceptual stimuli. This work demonstrates that asymmetry is usually recorded when relative baseline discriminability between the two dimensions is mismatched (e.g., Algom et al., 1996; Algom et al., 2022; Melara & Mounts, 1994; Pansky & Algom, 1999). Typically, the highly discriminable dimension leads to failure of selective attention to the less discriminable dimension but not the opposite (Melara & Mounts, 1993).

Garner Effects in Face Processing

The Garner paradigm has provided an effective tool for examining whether facial identity and other dimensions of face perception (such as, emotional expression and facial speech) are processed along a shared or parallel-route (see Algom & Fitousi, 2016; Fitousi & Wenger, 2013; Wang et al., 2013). Findings from several Garner studies provide evidence for the shared route hypothesis across dimensions of familiarity and sex (Ganel & Goshen-Gottstein, 2002) and identity and expression (Ganel & Goshen-Gottstein, 2004). In other studies, Garner interference was found to be asymmetrical. For example, interference of facial identity on emotional expression is found, but not the reverse (Schweinberger & Soukup, 1998; Wang et al., 2013). Similarly, sex classification has been found to influence emotion classification but not the opposite (Atkinson et al., 2005). Differences in discriminability of the two dimensions may explain the asymmetry recorded for facial identity processing and other facial dimensions when using the Garner Speeded Classification paradigm (Le Gal & Bruce, 2002; Wang et al., 2013). However, baseline discriminability may not explain all cases of asymmetry. For example, when discriminability is matched such that the speed of classifications of these two dimensions in the baseline condition is equal, asymmetry was upheld (Schweinberger & Soukup, 1998; Schweinberger et al., 1999).

There are also some inconsistencies in findings across studies, such that some studies have found symmetric Garner interference between two dimensions (e.g., identity and facial expression) (Fitousi & Wenger, 2013; Ganel & Goshen Gottstein, 2002) while others have recorded asymmetry (Schweinberger & Soukup, 1998; Schweinberger et al., 1999), or no interference at all (Wang et al., 2013). Disparities in findings may be attributed to variations in the stimuli used. Specifically, for example, whether relative baseline discriminability was controlled for, and whether the faces used were familiar or unfamiliar. With regards to the latter, face familiarity has been found to moderate the presence and strength of Garner interference in some studies (Ganel & Goshen-Gottstein, 2002, 2004; Kaufmann & Scheinberger, 2004). Thus, both these stimulus attributes will be considered in the present study.

The Present Study

Although age has been considered in studies assessing independence from emotional expression and has also been found to operate independently from sex and race (Fitousi, 2020), to date few studies have directly examined the independence of age and facial identity processing (Dagovitch & Ganel, 2010). The Garner speeded classification paradigm will be used to systematically investigate the independence of processing of age and facial identity. If age and facial identity share a processing mechanism, and are therefore *integral*, slower response times in the Filtering compared to the Baseline condition will be recorded. In contrast, null effects will be recorded across all experiments if facial identity and age processing are *separable*.

Given that existing theoretical models do not explicitly account for age perception (Bruce & Young, 1986; Burton et al., 1999; Haxby et al., 2000), in addition to the inconsistent findings on the interaction between age and facial identity processing in empirical studies (e.g., Chatterjee & Nakayama, 2012; Dagovitch & Ganel, 2010; Robertson & Burton, 2021), predicting a specific outcome is challenging. Findings from these experiments will serve to advance our understanding of age perception within existing theoretical frameworks of facial identity processing.

To this end, in Experiment 1, Garner paradigm will be used whereby participants will classify the age and facial identity of well-known famous faces across the two conditions (i.e., Baseline and Filtering). Given that previous studies have found familiarity effects on decision tasks of age (Bruyer et al., 2007) and other visually derived semantic codes, such as gender (Rossion, 2002) and ethnicity (Bruyer et al., 2004), as well as identity (Ganel & Goshen-Gottstein, 2002, 2004), an experiment with *unfamiliar* faces will be included. Thus, in Experiment 2, the design will be identical, but the facial identity stimuli will comprise lesser-known famous faces (e.g., international persons). The reason for using international individuals is because faces of unfamiliar people with known ages at different time points may be difficult to source.

Previous studies have reported that cues external to the facial features can facilitate categorisation by encouraging picture-based strategies (Goshen-Gottstein & Ganel, 2000). Given that the same stimuli are repeatedly presented to participants, any salient non-face marker, such as differences in face contour, could act as an artificial cue which may interfere with the categorisation task. Therefore, a similar approach to standardising stimuli (i.e., cropping faces into an oval shape) used in other studies (e.g., Wang et al., 2013) will be used here for Experiment 1 and 2 to discourage participants' from using such picture-based strategies. However, a criticism of the Garner speeded classification paradigm is the difficulty with generalising the findings to the normal face recognition system. This is because of the constrained nature of the cropped stimuli which is different to how we encounter faces in everyday life (see Burton et al., 2013). To address this, Experiment 3 will use multiple naturalistic images of famous identities comprising a high degree of within-person variability.

Experiment 1

Participants

To determine the sample size required for the main analysis, a 2 x 2 ANCOVA with one covariate, a power analysis was performed based on a study which examined the age classification paradigm for age and emotion categorisation. This study reported a significant interaction effect for condition (Filtering, Baseline) by categorisation (age, emotional expression) with an effect size of $\eta_p^2 = 0.28$ (extracted from Karnadewi & Lipp, 2011) whereby there was slower performance in the Filtering condition compared to the baseline condition such that varying age interfered with emotion but not the opposite. The power analysis computed using G*Power revealed a required sample size of 36 participants to achieve a power of 0.95 with an alpha level of 0.05. Therefore, 40 to 50 male and female participants were considered sufficient. All participants were required to have normal or corrected-to-normal vision. To minimise any cross-race influence on facial recognition performance (Bothwell et al., 1989; Meissner & Brigham, 2001) and age perception (Dehon & Brédart, 2001; Rhodes et al., 2009), only data from Caucasian participants is included. Participants recruited were over the age of 18, with no upper limit of age range. However, there is some evidence for an age bias in both facial recognition (Rhodes & Anastasi, 2012) and age perception (Davis & Attard-Johnson, 2022; Moyse & Brédart, 2012), therefore participant age was included as a covariate in the analysis. No other inclusion or exclusion criteria were specified.

In total 67 participants took part and fully completed Experiment 1. Of these, 11 were removed because they did not meet the criteria for ethnicity, and 1 because they had a % error rate of greater than 40%. This resulted in 55 remaining participants, 26 (16 male, M_{age} = 34, SD_{age} = 9.52, range = 20 to 55) completed the 'Age Classification' task, and 29 (16 male, 1 non-binary, M_{age} = 38, SD_{age} = 11.77, range = 19 to 66) completed the 'Facial Identification' task. Participants were randomly assigned to one of the two tasks. Participants who did not pass the initial prescreen or practice trials, or who did not see the task to completion, have not been reported here. Participants were recruited on Testable using the 'Verified Minds' participant pool between 12th February and 11th April 2024. All participants were compensated for their time. Ethical approval was obtained from the Institution's Ethics Board (Ref: 52251), and all participants provided consent to take part which was recorded electronically using Testable.

Stimuli

The faces used were standardised using the open-source GIMP (GNU Image Manipulation Programme) software. Therefore, all faces were cropped into an oval shape measuring around 8.5cm (width) x 12cm (height). Faces depicted a neutral facial expression and were presented in grey scale. To minimise variations across the dimensions other than those being examined (i.e., age and facial identity), all faces were male and Caucasian to maintain consistency for sex and race. Images that fit the criteria depicting faces in their 20's and in their 50's were sourced through an extensive google image search.

For Experiment 1, two models comprising famous faces (Matt Damon and Ewan McGregor) were used. These faces were chosen because both individuals are of a similar age (born in 1970 and 1971, respectively) and began their film career around the same period in the 1990's having performed in a similar number of films. Therefore, opportunity for exposure to both identities is roughly similar throughout their adult life. Nonetheless, to ensure that both identities are familiar to participants, a familiarity check will be performed prior to the experiment. Pilot data from 20 male and female participants (Mean age = 38 years, SD = 10.14) was collected for the four images to establish *perceived* age for the older ($McGregor_{Mean}$ = 44 and $Damon_{Mean}$ = 43 years) and younger faces ($McGregor_{Mean}$ = 30 and $Damon_{Mean}$ = 32 years). The image pairs for the young and older faces of the two identities were also compared for similarity using a 7 point-Likert scale (1 = "not at all similar" and 7 = "extremely similar"), and this revealed a low similarity score between Matt Damon and Ewan McGregor for both the older faces (M = 2.15, SD = 1.14) and younger faces (M = 2.35, SD = 1.39). Therefore, the images were considered sufficiently distinguishable.

Procedure

Participants completed a familiarity check by rating the faces (both older and younger versions) on their familiarity using a 5-point Likert scale. Only participants who provided a rating of 4 and 5 (highly familiar) were invited to complete the experiment, those who did not meet the criteria were screened out. Participants completed a familiarisation phase to ensure that they were clear of the identities (Model A vs Model B) and the two age categories (younger vs older). For this, there were 20 familiarisation trials to become acquainted with the four photographs and the appropriate response keys, thus participants were shown each photograph 5 times and given 5 seconds to respond. If an accuracy of at least 90% was not achieved, the familiarisation trials was repeated until reaching a maximum of three attempts. Participants who did not pass the initial accuracy by the third attempt were screened out.

Participants were randomly assigned to a task by the experiment software and depending on the condition assigned to, participants were instructed to classify the faces as either young or old, or Matt or Ewan using the keyboard. Participants were informed that speed and accuracy were both important, and to respond using the “Z” and “M” keys on a standard keyboard. Arrangement of the left and right response keys were balanced between participants, and the order was randomly generated by the experimental software. Each cycle consisted of three blocks: the two Baseline blocks and the Filtering block. In the Baseline blocks, participants judge the faces on one dimension (facial identity or age), while the other dimension is held constant (either Face A or Face B). Therefore, the baseline block consists of two parts, such that participants in the age classification group will judge age (young or old) for Model A and

Model B, separately. In the Filtering block, participants classify faces on one dimension while the other dimension varies. Therefore, participants in the age classification condition make age judgements for both Models A and B. Every participant completed three cycles. Within each cycle, the blocks were presented in a random order. The primary task comprised a total of 576 trials. Each trial started with a fixation cross for 300ms followed by a single face presented in the centre of the screen for 2500ms. A short break was provided every 48 trials (see Appendix for further details).

Results

Data Processing

Data was processed using R software (R Core Team, 2023) in the RStudio interface (Posit Team, 2024) using the R package 'tidyverse' (Wickham et al., 2019). Trials with a response time shorter than 150ms were removed from the analysis. Incorrect trials and trials without a response were removed. Percentage of trials removed are reported in Tables 1, 2, and 3. For the remaining trials, participant's mean RT aggregated across all Baseline blocks was calculated, producing a single value representing mean Baseline RT. The same calculation was performed for the Filtering condition. For the analysis, a 2 (condition: Baseline, Filtering) x 2 (classification task: age, facial identity) mixed-factorial ANCOVA, with participant age included as a covariate, was performed with partial eta square reported for effect sizes. Any significant interactions were followed-up with post-hoc *t*-tests adjusted for multiple comparisons using Bonferroni Correction.

This analysis was complemented with Bayes factors (BF) which can inform us about the null hypothesis which is important for assessing independence which will be operationalised as a null effect (Dienes, 2014). A Bayes Factors (BF_{10}) of greater than 3 represents substantial support for the alternative hypothesis, and smaller than 0.3 represents substantial support for the null hypothesis, and BF in between 0.3 and 3 represents only weak evidence (Jeffreys 1961; Wetzels et al., 2011). The traditional ANCOVA and Bayesian ANCOVA were both performed in JASP (version 0.18). The datasets generated and analysed during the current study are available in the Open Science Framework repository (URL: https://osf.io/85b2s/?view_only=d1da2126efff4932bd804124abe4277b).

Analysis

Within the Baseline conditions, the RTs for the two separate identities in the baseline condition of facial identification task (M_{diff} = 4ms), and for the two ages in the baseline condition of the age classification task, were comparable (M_{diff} = 11ms). Of primary interest are the RTs combined across identities and ages for the Filtering and Baseline conditions which are summarised in Table 1. The 2x2 mixed factorial ANCOVA revealed that age was not a significant covariate, F (1, 52) = 1.11, p = 0.295, ηp^2 = 0.021. There was no effect of classification task, F (1, 52) = 2.57, p = 0.115, ηp^2 = 0.05. A main effect of condition (Baseline v. Filtering) was present, F (1, 52) = 29.38, p < 0.001, ηp^2 = 0.36, such that response times in the Filtering condition were on average 62ms slower than the Baseline condition indicating the presence of Garner interference. The analysis also revealed an interaction between condition and classification task, F (1, 52) = 15.63, p < 0.001, ηp^2 = 0.23.

The data was also subjected to a 2 x 2 Bayesian ANCOVA. This revealed moderate evidence in favour of the null model for age as a covariate (BF_{10} = 0.75). The analysis revealed anecdotal evidence for the effect of classification task (BF_{10} = 1.36) and strong evidence for the interaction between classification and condition (BF_{inc} = 15.89). Strong evidence for the effect of condition was found (BF_{10} = 4.61 x 10^9), which supports the presence of Garner Inference.

[INSERT TABLE 1 HERE]

To unpack the interaction effect, four comparisons were performed. Post hoc *t*-tests were performed using an alpha level of 0.0125 (Bonferroni adjusted for four comparisons, 0.05/4 = 0.0125). When separated by age and identity classification tasks, paired *t*-tests revealed an increase in response time (40ms) from Baseline to Filtering in the age classification task, $t(25) = 5.25$, $p < 0.001$, $d = 1.03$ (BF_{10} = 1181, very strong). A similar pattern was found for the identity classification task (81ms), $t(28) = 8.69$, $p <$ 0.001, $d = 1.61$ (BF_{10} = 6.052 x 10^6, very strong). This demonstrates the presence of GI in both Age and Facial Identity tasks.

Independent samples *t*-tests comparing Baseline RTs for age and identity classification revealed that RTs for the Baseline condition were similar for both the age and identity classification tasks, $t(53) = 0.97$, $p = 0.336$, $d = 0.26$ (BF_{10} = 0.403, weak). Finally, the *t*-tests also demonstrated that the mean RTs for the Filtering trials were slightly lower for the identity classification task compared to the age classification task, but this did not reach the threshold for significance (alpha = 0.0125), $t(53) = 2.54$, $p =$ 0.014, $d = 0.26$ (BF_{10} = 3.70, weak).

Discussion

Experiment 1 demonstrated that varying age or identity of famous faces interfered with the classification of the other dimension, indicating the presence of Garner interference. Importantly, this interference was symmetrical, supporting the notion that facial identity and age processing are integral dimensions. The relationship between age and facial identity mirrors patterns recorded between identity and other facial attributed such as sex or race (e.g., Fitousi & Wenger, 2013; Ganel & Goshen-Gottstein, 2002). Additionally, the similar mean reaction times (RTs) for the Baseline tasks suggest that the two dimensions has equal relative baseline discriminability (RBD). RBD has been considered as a possible explanation for inconsistencies in Garner interference patterns across other studies (Schweinberger & Soukpu, 1998; Schweinberger et al., 1999). By matching RBD in this experiment, the likelihood of interference stemming from Baseline differences was reduced. Experiment 1 focussed on familiar faces using famous faces as stimuli. However, there are known differences in the processing of familiar and unfamiliar faces (see Burton, 2013; Johnston & Edmonds, 2009). Specifically, familiar face recognition benefits from an advantage in accuracy and speed in comparison to unfamiliar face recognition which is thought to rely more heavily on a less sophisticated picture-based strategy. Therefore, to determine whether the findings from Experiment 1 extend to unfamiliar face processing, a second experiment was conducted.

Experiment 2

The aim of Experiment 2 was to examine whether Garner interference recorded for famous faces extends to unfamiliar face processing. In Experiment 2, the design was identical but unfamiliar faces were used. Before completing the main task, participants learned the unfamiliar faces via a practice task, and therefore these faces became learned unfamiliar faces. There are two key differences between famous and unfamiliar faces that may influence the speed of classification in this study. First, use of internal facial features becomes more important for highly familiar faces, whereas external features are more relied upon for unfamiliar face identification (Ellis et al., 1979;

Kramer et al., 2018; Young et al., 1985). In this experiment, the faces are cropped to exclude external facial features, such as hair, which may result in a disadvantage for unfamiliar face classification. Second, drawing on the IAC model (Burton et al., 1999), highly familiar faces hold extra semantic information in the Personal Information Unit and Semantic Information Unit. This extra information may be accessed and serve to facilitate classification of age and identity. Indeed, previous studies have found a familiarity advantage, with faster response times for sex classifications of familiar faces compared to unfamiliar faces (Edmonds et al., 2013; Richards & Ellis, 2009). Consequently, it is expected that overall slower response times for the classification of age and identity for unfamiliar faces will be revealed.

However, predicting differences in presence or absence of Garner interference for highly familiar and unfamiliar faces is more complex. Most studies examining familiarity effects within the Garner paradigm focus on emotional expression as one dimension (e.g., Ganel & Goshen-Gottstein, 2002, 2004; Kaufmann & Scheinberger, 2004). Emotional expression is theoretically distinct from other non-identity face aspects such as sex and race due to its complex interconnected affective neural networks (e.g., Bruce & Young, 1986; Haxby et al., 2000). Nonetheless, familiarity has been shown to affect the presence and strength of Garner interference in studies examining emotional expression and identity (Ganel & Goshen-Gottstein, 2002, 2004; Kaufmann & Scheinberger, 2004), with stronger interference typically found for familiar faces. Therefore, it is possible that the interference between identity and age for unfamiliar faces is smaller or even absent.

Participants

In total 72 participants took part in the experiment, and 14 were removed because they did not meet the criteria for ethnicity. Error rates of remaining participants fell below 40% and therefore no additional participants were removed. This resulted in 58 remaining participants, 30 (11 male, 1 non-binary, $M_{age} = 35$, $SD_{age} = 14.75$, range = 18 to 71) completed the 'Age Classification' task, and 28 (11 male, 1 non-binary, $M_{age} = 33$, $SD_{age} = 11.90$, range = 19 to 59) completed the 'Facial Identification' task. Participants who did not pass the initial prescreen or practice trials, or who did not see the task to completion, have not been reported here. Participants were recruited on Testable using the 'Verified Minds' participant pool between 12th February and 11th April 2024 and had not participated in the previous experiment.

Stimuli and Procedure

For Experiment 2, two unfamiliar models were used. However, there is no existing database for images of unfamiliar faces taken decades apart to represent the young and old conditions required for this study. Therefore, Caucasian celebrities known outside of the United Kingdom were used instead. Several images sourced from an extensive google search were piloted for perceived age, and the two identities with ages most comparable to Experiment 2 were two European celebrities Fabrizio Faniello (a Maltese singer) and Johnny de Mol (a Dutch actor). Pilot data from 20 male and female participants was collected for the four images to establish *perceived* age for the older ($Faniello_{Mean} = 38$ and de $Mol_{Mean} = 42$ years) and younger faces ($Faniello_{Mean} = 28$ and de $Mol_{Mean} = 32$ years). The image pairs for the young and older faces of the two identities were also compared for similarity using a 7 point-Likert scale (1 = "not at all similar" and 7 = "extremely similar"), and this revealed a low similarity score between Fabrizio Faniello and Johnny de Mol for both the older faces (M = 2.3, SD = 1.08) and younger faces (M = 1.9, SD = 1.12). Therefore, the images were considered sufficiently distinguishable.

A familiarity check was performed before commencing with the experiment to ensure that participants were not familiar with the faces at the time of the experiment. Before commencing the study, participants completed a familiarity check on a pre-screen by rating the faces (both older and younger versions) on their familiarity using a 5-point Likert scale. Only participants who provided a rating 0 (not at all familiar) were invited to complete the experiment. The rest of the procedure was identical to Experiment 1, and to keep the instructions also identical, the unfamiliar models were assigned the false names 'Matt' and 'Ewan'.

Results

For the separate stimuli within the Baseline condition, the RTs for the two separate identities of facial identification task (M_{diff} = 4ms) were comparable, though slightly larger for the two ages in the baseline condition of the age classification task (M_{diff} = 21ms). The main analysis concerns the data for the RTs combined across identities and ages for the Filtering and Baseline conditions, these data are summarised in Table 2. A 2 (condition: Baseline, Filtering) x 2 (classification task: age, facial identity) mixed-factorial ANCOVA, with participant age included as a covariate, was performed. The analysis found a significant effect for the covariate of age, F (1, 55) = 12.33, $p < 0.001$, $\eta p^2 = 0.183$. Therefore, with age accounted for, the analysis found no effect of classification task on response time, F (1, 55) = 0.037, $p = 0.848$, $\eta p^2 = 0.0006$. A main effect of condition was found, F (1, 55) = 38.44, $p < 0.001$, $\eta p^2 = 0.411$, such that RTs were on average 96ms slower in the Filtering condition compared to Baseline. This means that Garner interference is present for both tasks. There was no significant interaction between condition and classification task, F (1, 55) = 3.85, p = 0.055, ηp^2 = 0.065.

The data was also subjected to a 2 x 2 Bayesian ANCOVA. This revealed strong evidence for the effect of age as a covariate (BF_{10} = 38.97). Once participant age is accounted for, the analysis revealed weak evidence for the effect of classification task (BF_{10} = 0.42) and for the interaction between classification and condition (BF_{inc} = 1.14). Strong evidence for the effect of condition was found (BF_{10} = 7.142 x 10^{17}), which supports the presence of Garner interference.

[INSERT TABLE 2 HERE]

Additional Analyses: Familiarity Comparison

To explore whether there were any differences in RTs for familiar and unfamiliar faces, separate 2 (familiarity: familiar, unfamiliar) x 2 (condition: Baseline, Filtering) frequentist and Bayesian ANCOVAs (participant age as a covariate) were performed for the identity and age classification tasks. This analysis is exploratory and was not specified in the pre-registered report.

For the identity classification task, this analysis revealed a main effect of condition, F (1, 54) = 39.98, p < 0.001, ηp^2 = 0.43, and as expected the RTs in the Filtering condition were overall slower than the Baseline condition. There was no main effect of familiarity, F (1, 54) = 0.04, p = 0.844, ηp^2 = 0.0007, and no interaction effect, F (1, 54) = 2.92, p = 0.093, ηp^2 = 0.051. The covariate of participant age was also not significant, F (1, 54) = 0.89, p = 0.350, ηp^2 = 0.016. The Bayesian equivalent analysis further supports this pattern demonstrating strong evidence for condition (BF_{10} = 1.38 x 10^{16}), and anecdotal evidence for the interaction (BF_{inc} = 1.65). In addition, BF revealed anecdotal evidence in favour of the null hypothesis for familiarity (BF_{10} = 0.31).

For the age classification task, this analysis revealed a main effect of condition, F (1, 53) = 21.82, p < 0.001, $\eta p^2 = 0.29$, and no main effect of familiarity, F (1, 53) = 2.71, p = 0.106, $\eta p^2 = 0.049$. In contrast to the identity task, an interaction between condition and familiarity was found, F (1, 53) = 15.28, p < 0.001, $\eta p^2 = 0.22$. To understand the interaction, independent t-tests comparing familiar and unfamiliar faces for Baseline and Filtering conditions were performed. These revealed faster RTs for the age classification of familiar faces (594ms) compared to the unfamiliar faces (672ms) in the Filtering condition, $t(54) = 2.21$, p = 0.032, d = 0.59, but the difference (554ms vs 588ms, respectively) was not significant for the Baseline condition, $t(54) = 1.04$, p = 0.302, d = 0.23. The Bayesian equivalent analysis further supports this pattern demonstrating extreme evidence for condition ($BF_{10} = 9.61 \times 10^{10}$), and strong evidence for the interaction (BF_{inc} = 67.21). In addition, BF revealed anecdotal evidence for familiarity (BF_{10} = 1.11). Furthermore, BF for the independent samples t-tests revealed anecdotal evidence in favour of the null hypothesis for the differences in RTs for the Baseline condition (BF_{10} = 0.42), and anecdotal evidence in favour of the alternative hypothesis for the Filtering condition (BF_{10} = 1.96).

Discussion

Taken together, these findings indicate that Garner interference is present, and symmetrical, and provide support for notion that facial identity and age processing are integral dimensions for both familiar and unfamiliar faces. The findings also show that discriminability for the two dimensions were of equal relative discriminability since the overall RT for the Baseline tasks were similar. RTs were compared directly for the familiar task in Experiment 1 and the unfamiliar task in Experiment 2. For the identity classification task, contrary to the prediction that familiarity would reveal an advantage in identity classification due to the more proficient use of internal facial cues and additional semantic information, responses were not significantly faster than for unfamiliar faces. However, an interaction was found for the age classification task. Specifically, in the Filtering condition, the ages of familiar faces were categorised faster than unfamiliar faces. Thus, although familiarity did not affect the presence of Garner interference, or the classification of identity, there was some detectable slowing down of age classification when unfamiliar face identities varied. It may be that the stored internal representation of familiar faces facilitates the processing of age to some extent by allowing participants disregard identity more rapidly and selectively attend to age. However, it should be noted that Bayes Factors only found anecdotal evidence for this effect, and there was also no strong evidence in preference of the null hypothesis (i.e., that familiarity does *not* affect the age of identity classification in the Garner paradigm). Importantly however, both highly familiar and unfamiliar faces show evidence for the shared route hypothesis with facial identity processing. Theoretical implications are discussed in more detail in the General Discussion.

Experiment 3

In recent years, researchers have argued against generalising experimental results from highly artificial and tightly constrained stimuli, for instance images with deliberate removal or manipulation of facial features of faces or other low-level image characteristics (see Burton, 2013). Furthermore, it is difficult to capture a true representation of a real person within a single image given large within-person variability in photos of the same face (Burton, 2013; Jenkins et al., 2011). To address this, Experiment 3 used an adaptation of the Garner speeded classification task to investigate whether effects are present with stimuli more likely to be encountered in daily life, specifically naturalistic images of the faces. Additionally, multiple images, as opposed to a single image, representing each identity was used.

Participants

In total 66 participants completed the experiment, and 17 were removed because they did not meet the criteria for ethnicity. Error rates of remaining participants fell below 40% and therefore no additional participants were removed. This resulted in 49 remaining participants, 24 (12 male, M_{age} = 36, SD_{age} = 9.14, range = 20 to 58) completed the 'Age Classification' task, and 25 (14 male, M_{age} = 38, SD_{age} = 9.7, range = 23 to 51) completed the 'Facial Identification' task. Participants who did not pass the initial prescreen or practice trials, or who did not see the task to completion, have not been reported here. Participants were recruited on Testable using the 'Verified Minds' participant pool between 11th April and 1st July 2024 and were independent to the participants in the previous two experiments.

Stimuli and Procedure

For Experiment 3, 20 ambient images for each of two famous identities from Experiment 1 were used with equal numbers (10 each) depicting younger faces (under 30) and older faces (over 40). The images of faces were sourced from a google search. Pilot data from 21 male and female participants was collected for all 40 images to establish *perceived* age for the older ($McGregor_{Mean}$ = 40 and $Damon_{Mean}$ = 43 years) and younger sets of faces ($McGregor_{Mean}$ = 25 and $Damon_{Mean}$ = 23 years). The images were adjusted to be of the same size, and although viewpoint and angle varied, all internal features were clearly visible. The background and face, as well as external features, were included. No other manipulations to the faces were made. Only participants who provide a rating of 4 and 5 (highly familiar) were invited to complete the experiment. The rest of the procedure was identical to the previous experiments, including the random presentation of the order of the images.

Results

Within the Baseline conditions, the RTs for the two separate identities in the baseline condition of facial identification task (M_{diff} = 28ms) were similar but larger than the previous two experiments, and comparable for the age classification task (M_{diff} = 13ms). Data of main interest for the RTs combined across identities and ages for the Filtering and Baseline conditions are summarised in Table 3. The 2x2 mixed factorial ANCOVA revealed that age was a significant covariate, F (1, 46) = 4.40, p = 0.04, ηp^2 = 0.009. With age accounted for, there was no main effect of condition (i.e., Filtering vs Baseline), F (1, 46) = 1.20, p = 0.16, ηp^2 = 0.042, indicating an absence of the Garner interference. There was also no interaction effect, F (1, 46) = 1.20, p = 0.16, ηp^2 = 0.042. The analysis found a main effect of classification task, F (1, 46) = 4.80, p = 0.033, ηp^2 = 0.095, such that response times on the age classification task were slightly slower than the identification task.

[INSERT TABLE 3]

The equivalent Bayesian analysis demonstrates weak evidence for the covariate of age (BF_{10} = 1.14). However, in contrast to the traditional ANCOVA, the analysis reveals strong evidence for the effect of condition (i.e., Filtering vs Baseline) (BF_{10} = 4505) indicating a slowing down of response time for Filtering trials providing evidence for the presence of Garner interference. There was weak evidence for the effect of classification task (BF_{10} =0.94) and for the interaction effect (BF_{10} = 1.51).

Discussion

Experiment 3 replicated the previous experiments with a few adjustments. First, naturalistic images of famous faces were used, and second, there were multiple images representing each identity and age as opposed to a single image of each. This was done to improve the generalisability of the findings by using images of faces more likely to be encountered in daily life (Burton, 2013). As with the previous experiments, results show that discriminability for the two dimensions were of equal relative discriminability since the overall RT for the Baseline tasks were similar.

A small Garner interference was present, but the effect was weaker than the previous experiments and although the frequentist analysis did not detect a significant effect, Bayes Factor revealed strong evidence of Garner interference for both the Age and ID classification tasks. On visual inspection of the descriptive data, the error rate is higher and Baseline RTs slower when compared to the previous experiments. This indicates that categorising highly variable faces either by age or identity required more cognitive resources or later stage of processing compared to the more highly controlled stimuli in the earlier experiments. This may be a result of processing additional visual information, such as colour, skin tone, and external features, which were not present in the cropped greyscale face stimuli used in Experiments 1 and 2. However, adding the second irrelevant dimension only resulted in a modest slowing down. These results are consistent with Burton's (2013) review on the limitations of generalising experimental results from highly artificial face stimuli to inform us about the normal face recognition system. The implications of these results for the parallel or shared route hypothesis are discussed in the next section.

General Discussion

Empirical work to date has examined whether facial identity and other aspects of the face, specifically, emotional expression, sex, and race, are processed through a single-route or in parallel to each other (e.g., Craig & Lipp, 2023; Ganel & Goshen-Gottstein, 2002; Wang et al., 2013). However, few studies have considered the interconnectedness of facial age with identity and non-identity aspects of the face (Alonso-Prieto et al., 2015; Fitousi, 2020). For the first time, the Garner Speeded Classification Task was employed to examine the relationship between facial identity processing and facial age perception. Using the standard Garner task, Experiments 1 and 2 revealed symmetrical Garner interference providing support for the shared route hypothesis. This was the case for both familiar and unfamiliar faces. Experiment 3 used an adaptation of the Garner task, by including multiple naturalistic images of two familiar identities as opposed to cropped single images per identity. In contrast to the first two experiments, only a weak Garner interference was present.

Shared or parallel-route for identity and age perception?

Experiments 1 and 2 recorded Garner interference for both the 'age judgement' condition and the 'identification judgement' condition. This is illustrated by a slowing down of response times when classifying a face across one dimension (e.g., age for the 'age judgement' condition) while the task-irrelevant dimension varies (e.g., identity for the 'age judgement' condition) (Filtering condition), as compared to when the task-irrelevant dimension is kept constant (Baseline condition). In accordance with the theoretical principles of the Garner Speeded Classification Task (Algom & Fitousi, 2016; Garner, 1976), the presence of Garner interference demonstrates a failure to selectively attend to a single stimulus dimension while ignoring the second. A difficulty in selectively attending to one dimension suggests that age and facial identity are integral dimensions and thus, processed along a shared route. Furthermore, the influence of the age and identity dimensions was symmetrical meaning that age interfered with the speed of facial identification in the facial identification condition and vice versa. These findings provide support for the hypothesis that age and identity are processed along a shared route.

These findings are similar to behavioural data from a small set of studies using the Garner Speeded Classification Task demonstrating that facial identity and another aspect of the face, sex, are also processed via a shared route (e.g., Ganel & Goshen-Gottstein, 2002; Ganel et al., 2002). However, intriguingly, age and sex have been found to operate in parallel to each other (see Fitousi, 2020). This suggests that although there is evidence that age and sex are processed independently, they are both individually integrated with facial identity. The same applies to the relationship between age, race, and facial identity. Specifically, there is evidence for the parallel-route hypothesis between race and age (Fitousi, 2020; Alonso-Prieto et al., 2015), and race and sex, (Fitousi, 2020), but support for a shared route between race and identity (Bruyer et al., 2004).

In contrast, findings on emotional expression and identity (Fitousi & Wenger, 2013; Schweinberger & Soukup, 1998; Wang et al., 2013), and age, sex, and race (see Atkinson et al., 2005; Karnadewi & Lipp, 2011; Le Gal & Bruce, 2002) point towards partial asymmetrical crosstalk indicated by asymmetric Garner interference. Specifically, in the studies where asymmetric interference is observed, emotion expression classification is influenced by all four of the above face processes (Baudin et al., 2002; Krebs et al., 2011; Schweinberger & Soukup, 1998; Wang et al., 2013), but there is no published record of asymmetry occurring in the opposite direction.

Additionally, a smaller set of experiments do record an absence of interference altogether (Craig & Lipp, 2023; Le Gal & Bruce, 2002), or symmetrical interference (Ganel & Goshen-Gottstein, 2004; Fitousi & Wenger, 2013).

Variations in methodology may account for the discrepant findings in previous work. For instance, a mismatch in relative baseline discriminability (RBD) (i.e., similarity of response times and accuracy in the Baseline conditions for the two single dimensions) (Melara & Mounts, 1993) has been found to produce asymmetric Garner interference in some studies (Wang et al., 2013) but not all (Schweinberger & Soukup, 1998; Schweinberger et al., 1999). Across all three experiments in the present study, reaction times and accuracy were similar across both age and identity dimensions in the Baseline conditions, which indicates that RBD was matched. Therefore, the possibility of obtaining asymmetrical interference because of mismatched relative baseline discriminability was reduced. Directly manipulating Baseline discriminability would be an important avenue for investigation to help shed light on its direct influence on the symmetry of Garner interference.

Although the evidence points towards the shared route hypothesis for age and identity perception, there is however, an important caveat to consider. Experiment 3, using unmodified naturalistic images of faces, found a weaker effect of Garner interference. This finding is consistent with Burton's (2013) proposal that conclusions drawn from experiments using faces stripped of their natural characteristics should not be extended to our natural face recognition system. A similar discrepancy between artificially manipulated face stimuli and naturalistic faces has been recorded in studies examining the independence of sex and identity. In these studies, a pattern of results supporting the shared route hypothesis was only found for cropped faces (Gottstein & Ganel, 2000), and not for intact faces (Bruce et al., 1987).

There is however also evidence to the contrary. Ganel et al (2002) experimentally manipulated intact and cropped images for the sex-identity dimensions in a Garner paradigm across four experiments and found Garner effects indicative of a shared route hypothesis when both faces removed of hair and intact faces were used. Interestingly, and similar to the findings from Experiment 3, Garner interference was absent when faces were both intact *and* multiple faces were used, in contrast to the single-cropped face paradigm. Garner interference has previously been recorded when using multiple images (e.g., Ganel & Goshen-Gottstein, 2002), thus it appears that there may be a potential interactive effect of using multiple images *and* intact ambient faces which weakens Garner interference. The reason for this is unclear. However, one potential explanation is that single-cropped image paradigms may encourage the use of picture-based heuristics, relying more on low-level image processing. This could account for the lack of differences in Garner interference observed with familiar and famous faces in Experiments 1 and 2, as the task may involve earlier visual processing stages rather than the deeper identity recognition or age perception. In contrast, using multiple varied and natural images might discourage low-level image processing, possibly requiring engagement with later processing stages, where age and identity could be processed more independently. This would suggest that findings from single-cropped image face stimuli may not represent our deeper face recognition system. However, there is an alternative consideration. If the single-cropped image approach does encourage low level picture-based strategy, then one might also predict that this would result in no Garner Interference at all, which is not the case of Experiments 1 and 2.

Taken together, the findings from this study suggest that age and identity are processed along a shared route, as evidenced by symmetrical Garner interference between the two dimensions. However, it is important to acknowledge the weaker Garner effect recorded for ambient face images and its potential implications. Different experimental designs, such as single versus multiple images, naturalistic versus modified, familiar versus unfamiliar, and the specific combination of dimensions being investigated across multiple studies (e.g., Fitousi, 2020; Schweinberger & Soukup, 1998; Ganel et al., 2002), add complexity to interpretation of these differences. Therefore, further systematic investigation is needed to fully clarify the conditions under which facial attributes are processed independently or along shared routes, potentially challenging the dichotomous perspective that these two processes are *either* separate or integrated.

Implications for face perception models

While acknowledging the complexities discussed in the previous section, the overall findings are in favour of the shared route hypothesis and add potentially new insights to existing theoretical models. Consider first the traditional Bruce and Young's (1986) theoretical model of face processing. Within this framework, age judgements are considered one of several *visually derived semantic* codes, information which can be inferred with relative accuracy from visual cues alone and are independent by-produces of face recognition, but not essential for face recognition. Thus, this model proposed a parallel-route between visually derived semantic codes (sex, race, and age) and facial identity recognition. Contrary to Bruce and Young's (1986) early model, sex, race, and age share a route with facial identification as evidenced by the symmetrical interference recorded in the current study and other Garner speeded classification tasks (Ganel &

Goshen-Gottstein, 2002; Ganel et al., 2002). Additionally, findings from this work, and previous behavioural experiments, suggest that these non-identity aspects of the face operate independently of *each other* (e.g., Alonso-Prieto et al., 2015; Fitousi, 2020). Bruce and Young (1986) also proposed that there was an *expression code* which was important for emotional expression recognition but not important for facial identification. In fact, the inconsistent partial interaction recorded between emotional expression classification and facial identity, age, race, and sex, supports the notion that emotional expression recognition is distinct from other non-identity aspects of the face (i.e., age, sex, and race).

Haxby et al's (2000) neurocognitive model of face perception expanded on Bruce and Young's model (1986) and separated these aspects of the face into two categories. The core system which involves the processing of invariant aspect of the face (such as, sex, race, and facial identity), and the extended system which processes changeable facial information (such as, emotional expression and facial speech patterns), for which the neural networks are independent. The extended system also includes person knowledge, such as biographical knowledge, episodic memories, and personality traits. Haxby et al (2000) emphasises the distinction between the invariant aspects of the faces which underlie recognition of identity, and changeable aspects of the face with facilitate social communication but not identity recognition. Facial age is not explicitly considered in Haxby's model leaving the question of whether age is to be considered as part of the core system or extended system unresolved. In this study, age interacts with facial identity in a similar way to known invariant facial aspects (e.g., sex), compared with changeable aspects (e.g., emotional expression), thus lending support to the possibility that age also forms part of the core system.

Furthermore, face familiarity did not influence the presence of Garner interference in these experiments, but familiarity *has* been found to influence Garner interference for emotional expression (Ganel et al., 2002) and speech pattern perception (Schweinberger & Soukup, 1998). In Haxby's model (Haxby & Gobbini, 2010), the extended system comprises emotional expression and speech pattern, as well as, aspects typically associated with facial familiarity, such as person knowledge. Therefore, emotional expression and person knowledge overlap within this extended system, which may explain why face familiarity affects Garner interference given that top-down semantic knowledge is present for familiar and not unfamiliar faces. In contrast, if age processing forms part of the core system, separate to the extended system where person knowledge resides, age processing is unaffected by face familiarity.

However, few studies have examined the effect of face familiarity directly with facial aspects other than speech patterns and emotional expression, therefore further research is needed for a clearer pattern to emerge and theoretical conclusions to be drawn. Additionally, although Haxby groups these facial characteristics into two systems (i.e., core and extended), how the different aspects of the face interact *within* these groups is not clear. The independence between sex, age, and race, and the interconnectedness of all three with facial identity, may point towards further nuance in the delineation between individual processes within the core system. However, this notion also requires further investigation.

Future directions

Despite the implications for a shared route between age and facial identity perception, there are additional questions to consider. First, in this study, age was measured using a dichotomous 'young' and 'old' classification to fit within the standard design of the Garner speeded classification paradigm. However, age perception can be measured using different approaches, such as numeric estimation by providing a precise estimate of the age, classification of faces as over or under the age of 18, and ordering and ranking faces by age. However, a recent study shows that different measurement approaches tap into different cognitive processes, and some processes may share more underlying cognitive mechanisms with facial identity than others (Attard-Johnson et al., 2024). This may explain discrepancies in studies relating to the behavioural interaction between age perception and facial identification. For instance, individuals with developmental prosopagnosia have been found to have unimpaired age processing pointing towards a parallel-route (see Bennetts et al., 2024; Chatterjee & Nakayama, 2012; Nunn et al., 2001), while other studies have found an interaction between age and face processing in neurotypical participants (see Dagovitch & Ganel, 2010; Robertson & Burton, 2021). These studies varied substantially in the methodological approach for measuring age perception. Therefore, although the simple age classification task described in this study offers evidence towards the shared route hypothesis, it is possible that there are nuances in this relationship based on the methodological approach used to measure age perception and further research is needed to tease this apart. Furthermore, although the actual age difference between the young and older categories was between 20 and 30 years, the age gap based on *perceived* ages was around 10 years. It may be informative to also compare identities across wider perceived age gaps, such as childhood and adult, revealing more drastic effects of aging on facial identity. However, such a manipulation may affect relative baseline discriminability as with a wider gap the age judgement classification task may become easier relative to the identity judgement task.

Second, relative baseline discriminability was matched across all three experiments, and it is not known whether a mismatch in discriminability along one of the stimulus dimensions would have replicated the asymmetric results obtained in some other face processing Garner studies (e.g., Schweinberger & Soukpu, 1998; Schweinberger et al., 1999; Wang et al., 2013). Though this may be a general limitation of the Garner speeded classification task potentially explaining some inconsistencies in symmetry of results, previous work having manipulated RBD directly have demonstrated that RBD cannot explain all cases of asymmetry (Schweinberger & Soukpu, 1998; Schweinberger et al., 1999). Future work could consider directly manipulating RBD for the identity-age relationship too.

Experiments 1 and 2 utilised artificially cropped images of the identities to minimise the possibility of non-face cues facilitating categorisation by encouraging picture-based strategies of classification (Goshen-Gottstein & Ganel, 2000). However, the constrained nature of the cropped stimuli is different to how we would encounter faces in everyday life (see Burton, 2013), and therefore may limit the generalisability of the findings in these experiments to normal face recognition. For this reason, a third experiment which was an adaptation of the Garner speeded classification paradigm was implemented. In Experiment 3, multiple ambient images of the identities were used. Although the same pattern emerged, there was a discrepancy between the frequentist analysis and Bayes Factor (BF). Specifically, strong evidence for the alternative hypothesis was found using BF but the frequentist analysis was non-significant. The reason for this discrepancy is unclear, however BF is more robust and reliable with smaller sample sizes (e.g., Hox et al., 2012; Lee & Song, 2004; van de Schoot et al., 2013). Furthermore, although the naturalistic images were used only for familiar faces, there is merit to applying this methodology to unfamiliar faces in the future. Unfamiliar face recognition is more vulnerable to the effects of naturalistic images and within-person variability than familiar face recognition (Burton, 2013), therefore the presence and strength of Garner interference may be affected for naturalistic images.

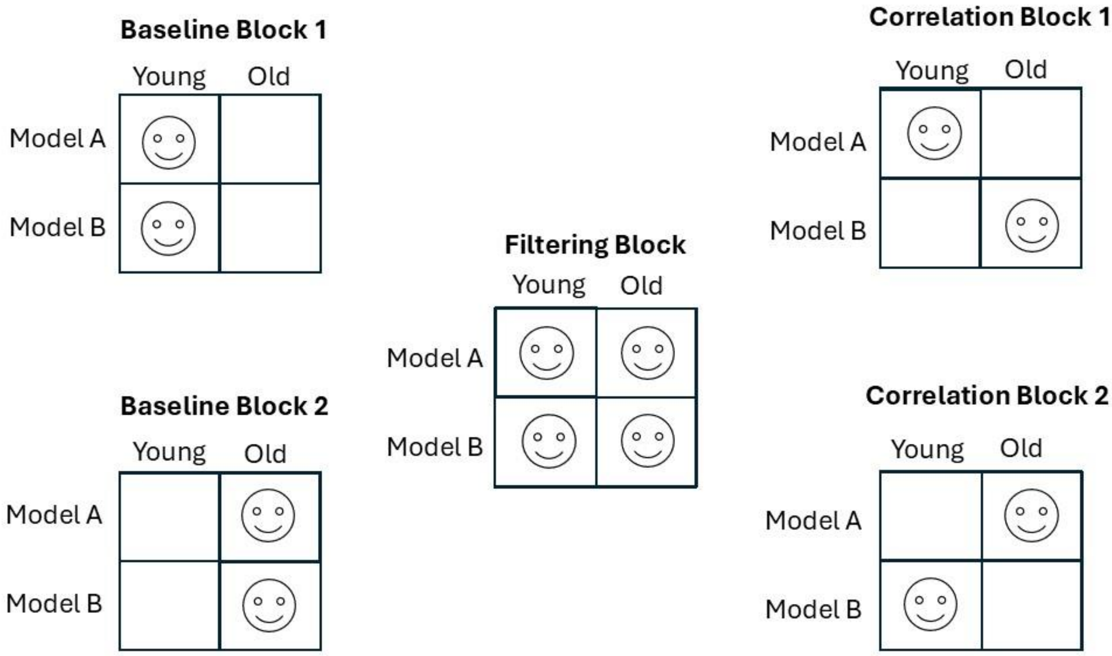
Finally, Garner Interference is a gold standard measure of dimensional interaction and selective attention, however, the precise mechanisms underlying the Garner Interference are not yet fully understood. It is beyond the scope of this paper to theorise about the underlying mechanisms of the paradigm itself. Recent novel accounts propose a binding account of facial dimensions (Fitousi, 2017a, b) and offer an alternative mechanism of binding explaining the Garner Interference (Fitousi, 2023). Specifically, the proposition is that facial attributes (e.g., identity, age, sex, emotion) interact in a binding process creating via the concept of a "face file", and that this binding can occur within- and across- independent processing routes, thus resulting in patterns that do not fully conform to the dual-route model. Additionally, sequential binding across consecutive trials can predict the magnitude of GI suggesting that the strength of feature integration influences the strength of interference recorded. For detailed explanation of this account see Fitousi (2017a, b; 2023). Future research may consider applying alternative and potentially converging methodologies to the examination identity and age. One such example is the conjoint measurement approach which has been used to study the effect of aging of faces on perceived gender (Fitousi, 2021) finding that observers integrate gender and age information.

Conclusion

Using the Garner speeded classification task, the current study provides evidence for a shared route between facial age and identity processing. Considering the body of work using Garner interference, it seems that emotional expression relates more inconsistently with facial identity, age, sex, and race. While studies examining the identity-age, identity-sex, and identity-race interconnectedness thus far appear to support the shared route hypothesis. In contrast, studies examining the interconnectedness of age-sex, age-race, sex-race, age-race, suggest that these processes occur independently via a parallel-route. However, the number of studies examining emotional expression far outnumbers other non-identity aspects of the face, therefore further research is needed to build a clearer picture of the interconnectedness between the different non-identity aspects of the face and facial identity. Furthermore, differences emerged between highly controlled and naturalistic face stimuli which highlight the complexities with interpretations across studies employing different methodologies.

Appendix

Figure 1. Graphical illustration depicting which stimulus combinations would be presented for each of the Baseline, Filtering, and Correlation Blocks. The two dimensions in the example below are Facial Identity (Model A/Model B) and Facial Age (Young/Old), and the task is to classify facial identity while ignoring the dimension of age. The figure is an adaptation based on Algom and Fitousi's (2016) graphical representation of the Garner paradigm. *Note that in the present study only the Baseline and Filtering Blocks were used.*



Appendix

Table 1

Mean RTs (in ms) and proportion of error (%) indicating average percentage of trials excluded for the Baseline and Filtering trials in Experiment 1 (Famous Faces). SD = standard deviation.

	Age				Facial ID			
	RT (ms)		Error (%)		RT(ms)		Error (%)	
	M	SD	M	SD	M	SD	M	SD
Baseline	554.10	68.58	3.25	2.53	575.75	93.17	2.56	2.18
Filtering	594.53	72.40	5.21	3.61	656.87	104.30	4.36	3.00
Garner Interferences	-40.43				-81.12			

Table 2

Mean RTs (in ms) and proportion of error (%) indicating average percentage of trials excluded for the Baseline and Filtering trials in Experiment 2 (Unfamiliar Faces). SD = standard deviation.

	Age				**Facial ID**			
	RT (ms)		**Error (%)**		**RT(ms)**		**Error (%)**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Baseline	588.48	155.63	3.47	3.16	561.79	84.48	4.13	4.14
Filtering	672.36	672.36	5.00	4.89	671.05	106.11	5.62	5.37
Garner Interferences	-83.88				-109.26			

Table 3

Mean RTs (in ms) and proportion of error (%) indicating average percentage of trials excluded for the Baseline and Filtering trials in Experiment 3 (Ambient Famous Faces). SD = standard deviation.

	Age				**Facial ID**			
	RT (ms)		**Error (%)**		**RT(ms)**		**Error (%)**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Baseline	699.91	79.42	7.43	5.64	667.89	73.37	7.23	4.61
Filtering	728.79	88.36	7.33	4.73	681.44	79.34	5.12	5.12
Garner Interferences	-28.88				-13.55			

References

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition, 24,* 557-572. https://doi.org/10.3758/BF03201083

Algom, D., & Fitousi, D. (2016). Half a century of research on Garner interference and the separability-integrality distinction. *Psychological Bulletin, 142,* 1352-1383. https://doi.org/10.1037/bul0000072

Algom, D., & Fitousi, D., & Chajut, E. (2022). Can the Stroop effect serve as the gold standard of conflict monitoring and control? A conceptual critique. *Memory & Cognition, 50,* 883-897. https://doi.org/10.3758/s13421-021-01251-5

Alonso-Prieto, E., Oruç, I., Rubino, C., Zhu, M., Handy, T., & Barton, J. J. (2015). Interactions between the perception of age and ethnicity in faces: an event-related potential study. *Cognitive Neuropsychology*, *32*, 368-384.

Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between sex and emotion in face perception. *Perception & Psychophysics*, *67*, 1199-1213. https://doi.org/ 10.3758/BF03193553

Attard-Johnson, J., Dark, O., Murray, E. & Bate, S. (in press). Different Facets of Age Perception in People with Developmental Prosopagnosia and "Super-Recognisers. *Cognitive Research: Principles and Implications*.

Bate, S., Bennetts, R. J., Gregory, N., Tree, J. J., Murray, E., Adams, A., ... & Banissy, M. J. (2019). Objective patterns of face recognition deficits in 165 adults with selfreported developmental prosopagnosia. *Brain Sciences*, *9*, 133. https://doi.org/10.3390/brainsci9060133

Baudouin, J. Y., Martin, F., Tiberghien, G., Verlut, I., & Franck, N. (2002). Selective attention to facial emotion and identity in schizophrenia. *Neuropsychologia*, *40*, 503-511.

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioural findings. *Journal of Cognitive Neuroscience, 17,* 1130-1149. https://doi.org/10.1162/0898929054475154

Bennetts, R. J., Gregory, N. J., & Bate, S. (2024). Both identity and non-identity face perception tasks predict developmental prosopagnosia and face recognition ability. *Scientific Reports*, *14*, 6626.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin*, *15*, 19-25. https://doi.org/10.1177/0146167289151002

Bruce, V., Ellis, H. D., Gibling, F., & Young, A. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology*, *41*, 510.

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

Bruyer, R., Leclére, S., & Quinet, P. (2004). Ethnic categorisation of faces is not independent of face identity. *Perception, 33*, 169–179. https://doi.org/10.1068/p5094

Bruyer, R., Mejias, S., & Doublet, S. (2007). Effect of face familiarity on age decision. *Acta Psychologica, 124,* 159-176. https://doi.org/10.1016/j.actpsy.2006.03.001

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The important of variability. *The Quarterly Journal of Experimental Psychology, 66,* 1467-1485. https://doi.org/10.1080/17470218.800125

Burton, A. M., Bruce, V., & Hancock, P. J. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, *23*, 1-31. https://doi.org/10.1207/s15516709cog2301_1

Cant, J. S., Large, M. E., McCall, L., & Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception*, *37*, 57-78.

Chatterjee, G., & Nakayama, K. (2012). Normal facial age and gender perception in developmental prosopagnosia. *Cognitive Neuropsychology*, *29*, 482-502. https://doi.org/10.1080/02643294.2012.756809

Craig, B. M., & Lipp, O. V. (2023). Bodily cues of sex and emotion can interact symmetrically: Evidence from simple categorization and the Garner paradigm. *Emotion*, *23*, 2385-2398. https://doi.org/10.1037/emo0001243

Dagovitch, Y., & Ganel, T. (2010). Effects of facial identity on age judgements: Evidence from repetition priming. *Experimental Psychology, 57,* 390-397. https://doi.org/10.1027/1618-3169/a000047

Davis, H., & Attard-Johnson, J. (2022). Your ID, please? The effect of facemasks and makeup on perceptions of age of young adult female faces. *Applied Cognitive Psychology*, *36*, 453-459. https://doi.org/10.1002/acp.3923

Dehon, H., & Brédart, S. (2001). An 'other-race' effect in age estimation from faces. *Perception*, *30*, 1107-1113. https://doi.org/10.1068/p3122

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Edmonds, A. J., Johnston, R. A., & Clutterbuck, R. (2013). The influence of familiarity on sex decisions. *Visual Cognition*, *21*, 853-875.

Eidels, A., Townsend, J. T., & Algom, D. (2010). Comparing perception of Stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition*, *114*, 129-150.

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

Fitousi, D. (2017a). Binding sex, age, and race in unfamiliar faces: The formation of "face files". *Journal of Experimental Social Psychology*, *71*, 1-15.

Fitousi, D. (2017b). What's in a "face file"? Feature binding with facial identity, emotion, and gaze direction. *Psychological Research*, *81*, 777-794.

Fitousi, D. (2020). Evaluating the independence of age, sex, and race in judgment of faces. *Cognition*, *202*, 104333. https://doi.org/10.1016/j.cognition.2020.104333

Fitousi, D. (2021). How facial aging affects perceived gender: Insights from maximum likelihood conjoint measurement. *Journal of Vision*, *21*, 12-12.

Fitousi, D. (2023). Dimensional interaction is driven by feature integration: A binding theory of Garner interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49,* 1881-1908.

Fitousi, D., & Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 133-155. https://doi.org/10.1037/a0028001

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science*, *19*, 207-208. https://doi.org/10.3758/BF03328777

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrality of sex and identity of faces: further evidence for the single-route hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 854. https://doi.org/10.1037/0096-1523.28.4.854

Ganel, T., & Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 583-597. https://doi.org/10.1037/0096-1523.30.3.583

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*, 98-123. https://doi.org/10.1016/0010-0285(76)90006-2

Garner, W. R. (1977). The effect of absolute size on the separability of the dimensions of size and brightness. *Bulletin of the Psychonomic Society*, *9*, 380-382.

Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgement task: evidence for a common route for the processing of sex and identity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1198-1214. https://doi.org/10.1037/0278-7393.26.5.1198

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223-233. https://doi.org/10.1016/S1364-6613(00)01482-0

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3,* 80-84. https://doi.org/10.1038/71152

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*, 87-93.

Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313-323.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*, 577-596. https://doi.org/10.1080/09658210902976969

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A model in human extrastriate cortex specialised for face perception. *The Journal of Neuroscience, 17,* 4302-4311. https://doi.org/0270-6474/97/174302

Karnadewi, F., & Lipp, O. V. (2011). The processing of invariant and variant face cues in the Garner Paradigm. *Emotion*, *11*, 563. https://doi.org/10.1037/a0021333

Kaufmann, J. M., & Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception,* 33, 399-408. https://doi.org/ 10.1068/p5083

Kramer, R. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46-58.

Krebs, J. F., Biswas, A., Pascalis, O., Kamp-Becker, I., Remschmidt, H., & Schwarzer, G. (2011). Face processing in children with autism spectrum disorder: independent or interactive processing of facial identity and facial expression? *Journal of Autism and Developmental Disorders*, *41*, 796-804.

Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*, 653-686.

Le Gal, P. M., & Bruce, V. (2002). Evaluating the independence of sex and expression in judgments of faces. *Perception & Psychophysics*, *64*, 230-243. https://doi.org/10.3758/BF03195789

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition, 61,* 139-158. https://doi.org/ 10.1016/j.bandc.2005.11.005

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3. https://doi.org/10.1037/1076-8971.7.1.3

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, *48*, 169-178.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21,* 627-645. https://doi.org/10.3758/BF03197195

Melara, R. D., & Mounts, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception and Psychophysics, 56,* 73-90. https://doi.org/10.3758/bf03211692

Mileva, M., Young, A. W., Jenkins, R., & Burton, A. M. (2020). Facial identity across the lifespan. *Cognitive Psychology*, *116*, 101260. https://doi.org/10.1016/j.cogpsych.2019.101260

Moyse, E., & Brédart, S. (2012). An own-age bias in age estimation of faces. *European Review of Applied Psychology*, *62*, 3-7. https://doi.org/10.1016/j.erap.2011.12.002

Nunn, J. A., Postma, P., & Pearson, R. (2001). Developmental prosopagnosia: Should it be taken at face value? *Neurocase, 7,* 15-27. https://doi.org/10.1093/neucas/7.1.15

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology, 25,* 39-58. https://doi.org/10.1037/0096-1523.25.1.39

Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of facial contrast decrease with age and are cues for age perception. *PloS one*, *8*, e57985. https://doi.org/10.1371/journal.pone.0057985

Posit team (2024). *RStudio: Integrated Development Environment for R.* Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology, 23*, 1-12. https://doi.org/10.1002/acp.1442

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*, 146. https://doi.org/10.1037/a0025750

Richards, R. M., & Ellis, A. W. (2009). Identity, gender, and the role of age of acquisition in face processing. *Psicológica*, *30*, 155-179.

Robertson, D. J., & Burton, A. M. (2021). Checking ID-cards for the sale of restricted goods: Age decisions bias face decisions. *Applied Cognitive Psychology*, *35*, 71-81. https://doi.org/10.1002/acp.3739

Rosenthal, G., & Avidan, G. (2018). A possible neuronal account for the behavioural heterogeneity in congenital prosopagnosia. *Cognitive Neuropsychology, 35,* 74-77. https://doi.org/10.1080/02643294.2017.1417248

Rossion, B. (2002). Is sex categorization from faces really parallel to face recognition? *Visual Cognition, 9,* 1003–1020. https://doi.org/10.1080/13506280143000485

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, *35*, 749-759. https://doi.org/10.1068/p5490

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1748. https://doi.org/10.1037/0096-1523.24.6.1748

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics*, *61*, 1102-1115. https://doi.org/10.3758/BF03207617

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrality of sex and identity of faces: further evidence for the single-route hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 854.

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important than how You Test. *Journal of Cognition*, *6,* 1–17. DOI: https://doi.org/10.5334/ joc.259

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

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 842-860.

Wang, Y., Fu, X., Johnston, R. A., & Yan, Z. (2013). Discriminability effect on Garner interference: evidence from recognition of facial identity and expression. *Frontiers in Psychology*, *4*, 943. https://doi.org/10.3389/fpsyg.2013.00943

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*, 291-298. https://doi.org/10.1177/1745691611406923

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software, 4,* 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

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

APPENDIX

Pre-registered Design and Analytical Approach

The experimental design and analyses approach were pre-registered and are reported in this section. Participants will complete two blocks (Baseline, Filtering) of the Garner paradigm (Garner, 1974) but will only complete one of the classification tasks (i.e., judgement of age or identity). The Garner paradigm has traditionally also included a correlated block; however, this is of less theoretical significance, and as with previous studies (Atkinson et al., 2005; Karnadewi & Lipp, 2011; Wang et al., 2013) will be omitted for simplicity. The between-subjects approach for the classification task follows previous work and reduces the possibility of carry-over in attention and strategy when switching classification task (Karnadewi & Lipp, 2011; Wang et al., 2013). In the Baseline block, participants will judge the faces on one dimension (facial identity or age), while the other dimension is held constant (either Face A or Face B). Therefore, the Baseline block will consist of two parts, such that participants in the age classification group will judge age (young or old) for Model A and Model B, separately. In total there will be 48 trials per part, with a total of 96 trials for the baseline block. In the Filtering block, participants will judge the faces on one dimension while the other dimension will vary. Therefore, participants in the age classification condition will make age judgements for both Models A and B. The Filtering block will have a total of 96 trials. One cycle of Baseline and Filtering blocks will comprise a total of 192 trials, and participants will complete a total of three cycles. The tasks will be administered online using Testable using the *Verified Minds* participant pool. Comparisons between lab-based and online data collection for cognitive tasks, even with complex paradigms, reveal acceptable levels of loss of data quality for web-testing and support the use of web-testing (Uittenhove et all., 2023).

To determine the sample size required for the main analysis, a 2 x 2 ANCOVA with one covariate, a power analysis was performed based on a study which examined the age classification paradigm for age and emotion categorisation. This study reported a significant interaction effect for condition (Filtering, Baseline) by categorisation (age, emotional expression) with an effect size of $\eta_p^2 = 0.28$ (extracted from Karnadewi & Lipp, 2011) whereby there was slower performance in the Filtering condition compared to the baseline condition such that varying age interfered with emotion but not the opposite. The power analysis computed using G*Power revealed a required sample size of 36 participants to achieve a power of 0.95 with an alpha level of 0.05. Therefore, 40 to 50 male and female participants for each experiment will be recruited resulting in an estimated total of 150 participants across all three experiments. All participants will have normal or corrected-to-normal vision. To minimise any cross-race influence on facial recognition performance (Bothwell et al., 1989; Meissner & Brigham, 2001) and age perception (Dehon & Brédart, 2001; Rhodes et al., 2009), all participants will be Caucasian. Participants over the age of 18 will be invited to take part, there will be no other limits on the participant age range. However, there is some evidence for an age bias in both facial recognition (Rhodes & Anastasi, 2012) and age perception (Davis & Attard-Johnson, 2022; Moyse & Brédart, 2012), therefore participant age will be included as a covariate in the analysis.

Using RStudio, each participant's mean RTs and percentage error will be computed for the baseline condition and the Filtering condition. Trials with a response time shorter than 150ms will be removed from the analysis. Incorrect trials and trials without a response will also be removed. Percentage of trials removed will be reported. For the analysis, a 2 (condition: baseline, Filtering) x 2 (classification task: age, facial identity) mixed-factorial ANCOVA, with participant age included as a covariate, will be performed. Partial eta square will be reported for effect sizes. Any significant interactions will be followed-up with post-hoc *t*-tests adjusted for multiple comparisons using Bonferroni Correction. This analysis will be complemented with Bayes factors (BF) which can inform us about the null hypothesis which is important for assessing independence which will be operationalised as a null effect (Dienes, 2014). A Bayes Factors (BF_{10}) of greater than 3 will represent substantial support for the alternative hypothesis, and smaller than 0.3 will represent substantial support for the null hypothesis, and BF in between 0.3 and 3 represents only weak evidence (Jeffreys 1961; Wetzels et al., 2011). The traditional ANCOVA and Bayesian ANCOVA will both be performed in JASP (version 0.18). The datasets generated and analysed during the current study are available in the Open Science Framework repository (URL: https://osf.io/85b2s/?view_only=d1da2126efff4932bd804124abe4277b).