Received: 3 December 2019 | Revised: 16 July 2020 | Accepted: 16 September 2020
DOI: 10.1002/acp.3744

RESEARCH ARTICLE

WILEY

How old was he? Disguises, age, and race impact upon age estimation accuracy

Craig Thorley

Department of Psychology, James Cook University, Townsville, Australia

Correspondence
Craig Thorley, Department of Psychology, James Cook University, Townsville, 4814, Australia.
Email: craig.thorley@jcu.edu.au

Summary

There are many situations in which a person must accurately estimate a stranger's age. For example, a salesperson must accurately estimate the age of a stranger who wishes to purchase age-restricted goods. Information from a stranger's eyes and hair can indicate their likely age. Here, two experiments examined whether adult participants' accuracy when estimating strangers' ages is reduced when the strangers' eyes are disguised by sunglasses and/or hair is disguised by a hat. The strangers' age and race also varied, and participants' social contact with the other age/race groups measured. At best, estimations of undisguised strangers' ages were inaccurate by an average of 5.10 years. Accuracy decreased when the strangers' eyes, but not hair, were disguised and when they came from other age/race groups. Accuracy when estimating other age/race group members' ages and social contact with their members were unrelated. The theoretical and applied implications of these findings are discussed.

KEYWORDS

age estimation, eyewitness memory, own-age bias, own-race bias, social contact

1 | INTRODUCTION

There are many situations in which people must accurately estimate strangers' ages. For example, salespeople must accurately estimate the age of unfamiliar customers who wish to purchase age-restricted goods or receive age-related discounts. In some situations, such as when bar staff are selling alcohol to customers at outdoor music festivals, the customers' eyes can be disguised by sunglasses and/or hair disguised by a hat. Importantly, these disguises could make it harder to estimate strangers' ages as information from their eye region and hair can indicate their likely age (Rhodes, 2009). The primary aim of the current research was to examine whether adult participants' accuracy when estimating strangers' ages is reduced if the strangers' eyes are disguised by sunglasses and/or hair is disguised by a hat. A secondary aim was to examine whether estimations of strangers' ages are any less accurate when strangers come from other age or racial groups, and whether accuracy increases as social contact with members of those groups increases.

1.1. | Age estimation accuracy and bias

Age estimation accuracy is often studied by presenting participants with passport-style photographs of strangers and asking them to estimate the strangers' ages in years. Age estimation accuracy and/or age estimation bias are then calculated. Age estimation accuracy is an absolute measure of how accurate age estimations are. It is calculated by subtracting each age estimation (e.g., 25 years) from each stranger's chronological age (e.g., 30 years) and working out the mean absolute error or MAE (e.g., 5 years). Across the literature, age estimations are often inaccurate by an average of close to 5 years (see Moyse, 2014, for a review). For example, Voelkle, Ebner, Lindenberger, and Riediger (2012) had young adults estimate the age of 19–80-year-olds and found the average MAE was 5.91 years. Similar averages have been found when participants estimate the age of strangers in videos (Amilon, Weijer, & Schötz, 2007) and when sexual assault victims estimate the age of stranger offenders during police interviews (Thorley, Almond, Gregory, McAlonan, & McLoughlin, 2018).

When calculating MAE, any negative values from underestimations are treated as positive values. This stops underestimations (e.g., −10 years) and overestimations (e.g., +8 years) from cancelling each other out during the averaging process, which would produce overly conservative MAE scores (e.g., −1 year). MAE scores do not, therefore, indicate whether participants had an age estimation bias, which is a tendency to underestimate or overestimate strangers' ages. Age estimation bias is calculated in a similar way to MAE, except underestimations are treated as negative values during the averaging process. To date, few systematic biases have been identified but one consistent finding, relevant to the current research, is that participants often overestimate the age of young adults and underestimate the age of elderly adults, with estimations towards middle-aged adults falling in-between and being least biased (e.g., Henss, 1991; Short, Mondloch, deJong, & Chan, 2019; Sörqvist & Eriksson, 2007; Vestlund, Langeborg, Sörqvist, & Eriksson, 2009; Voelkle et al., 2012; Willner & Rowe, 2001).

1.2. | Faces and age estimations

Facial features change in significant, but predictable, ways with age and the extent of these changes can indicate a person's likely age. These changes include the eyes getting smaller and sinking deeper into their orbits, the ears and nose becoming elongated due to the effects of gravity, and the lips becoming thinner. Importantly, the size and shape of these features are used to estimate strangers' ages (see Rhodes, 2009, for a review). To demonstrate this, George and Hole (1998) obtained photographs of the same individuals at two different ages and digitally swapped their eyes, nose, and mouth. Adding the older features to the younger faces increased their perceived age by nearly 40%, whereas adding the younger features to the older faces decreased their perceived age by nearly 33%.

Facial skin also changes in significant, but predictable, ways with age. For example, the skin develops wrinkles, which gradually increase in number, length, and depth. The skin also becomes thinner, paler, and saggier. The extent of these changes is also used to estimate a strangers' age. For example, it has been shown that a person's skin texture (e.g., number of wrinkles) and colour (e.g., how pale it is) both predict their perceived age (Gunn et al., 2009; Mayes et al., 2010; Merinville, Grennan, Gillbro, Mathieu, & Mavon, 2015; Nkengne et al., 2008). Moreover, digitally reducing a person's wrinkles, colour changes, and sagging can reduce their perceived age (Burt & Perrett, 1995; Fink, Grammer, & Matts, 2006; Fink & Matts, 2008; George & Hole, 1995; Porcheron, Latreille, Jdid, Tschachler, & Morizot, 2014; Samson, Fink, Matts, Dawes, & Weitz, 2010). Similarly, digitally increasing the number of wrinkles a person has can increase their perceived age (Aznar-Casanova, Torro-Alves, & Fukusima, 2010).

Several studies show the strongest predictors of a stranger's perceived age come from their eye region, such as the number of wrinkles around the eyes (El Dib & Onsi, 2011; Lanitis, 2002; Merinville et al., 2015; Nkengne et al., 2008). Disguising the eye region could, therefore, make it harder to accurately estimate a stranger's age. Jones and Smith (1984) appear to be the only researchers to have investigated this possibility. They had pre-school children rank photographs of child and adult strangers by age. The strangers' faces were either undisguised, had their eye region disguised by opaque tape, or had their facial outline (including head shape and hair) disguised by opaque tape. The age rankings were least accurate when the strangers' eye region was disguised, confirming the importance of that region in estimating a stranger's age.

1.3. | Hair and age estimations

As people age, their hair typically becomes greyer, thinner, and balding can occur. These changes are also used to estimate strangers' ages, although overestimations can occur if changes occur prematurely (e.g., premature balding). For example, Bulpitt, Markowe, and Shipley (2001) and Gunn et al. (2009) found that as the amount of grey hair present in 37–81-year-olds increased, the older they were perceived to be. Increased baldness in 30–58-year-old males and increased hair thinning in 45–81-year-old females also increased perceived age (Bulpitt et al., 2001; Butler, Pryor, & Grieder, 1998; Gunn et al., 2009). Rexbye et al. (2005), however, only found a weak correlation between baldness levels in 70–91-year-old males and their perceived age, which is perhaps unsurprising as balding/baldness is associated with men in that age range.

Only a few studies have examined whether disguising a stranger's hair makes it harder to accurately estimate their age. In those studies, strangers had their entire head shape disguised, meaning only their internal facial features were visible. In the Jones and Smith (1984) study discussed earlier, children's ability to rank strangers by age was no worse when the strangers' head shape and hair were disguised, relative to when they were undisguised (see also George, Hole, & Scaife, 2000). Similarly, George and Hole (1995) found adults' estimations of 5–70-year-old strangers' ages were no worse when the strangers' head shape and hair were disguised. Combined, the findings discussed so far all suggest that when people estimate strangers' ages, their age estimations may be more influenced by the strangers' facial features and skin than the strangers' hair.

1.4. | Age, race, and age estimations

How accurately a person can estimate a stranger's age may be influenced by the stranger's actual age. Several studies show that young adults have an own-age advantage when estimating strangers' ages, meaning they are better at estimating the age of strangers from their own age group than other age groups (Anastasi & Rhodes, 2006; George & Hole, 1995; Klugman, 1947; Moyse & Brédart, 2012; Short et al., 2019; Thorley et al., 2018; Voelkle et al., 2012), even if they overestimate the age of young adults and underestimate the age of older adults (e.g., Henss, 1991; Short et al., 2019; Sörqvist & Eriksson, 2007; Vestlund et al., 2009; Voelkle et al., 2012; Willner & Rowe, 2001). Why the own-age-advantage occurs is unclear but one potential explanation is that young adults have more social contact with members of their own age group, so have developed perceptual expertise in estimating the age of other young adults (George & Hole, 1995; Moyse & Brédart, 2012). If true, young adults who have regular social contact with older adults should have developed perceptual expertise in estimating their ages and have less of an own-age advantage. None of the above studies, however, measured participants' social contact with other age groups, so this perceptual-expertise explanation is untested.

Dehon and Brédart (2001) also found evidence of an own-race advantage in age estimation accuracy, whereby their young to middle-aged White participants were better at estimating the age of young to middle-aged White strangers than young to middle-aged Black strangers. Moreover, overestimations occurred regardless of race. Dehon and Brédart felt this own-race advantage may also be consistent with a perceptual-expertise explanation, as their White participants would have likely had more social contact with other White people. If true, White people who have regular social contact with Black people should have developed perceptual expertise in estimating their ages and have less of an own-race advantage. Dehon and Brédart did not, however, measure other-race contact and this perceptual-expertise explanation remains untested.

1.5. | Aims and hypotheses

Here, two experiments were conducted. Their primary aim was to determine whether adult participants' accuracy when estimating adult strangers' ages is reduced when the strangers' eye region is disguised by sunglasses and/or hair is disguised by a hat, relative to when neither is disguised. Past research shows information from a stranger's eye region is used to estimate their age (e.g., Nkengne et al., 2008) and that children are worse at estimating strangers' ages when the strangers' eye region is disguised (Jones & Smith, 1984). Consistent with these findings, it was predicted that participants in the current experiments would be worse at estimating strangers' ages when those strangers have their eye region disguised by sunglasses. Past research also suggests that information from a stranger's hair can be used estimate their age (e.g., Bulpitt et al., 2001) but that estimations of strangers' ages are no worse when their entire head shape and hair is disguised (e.g., George & Hole, 1995). It is, therefore, possible that participants in the current experiments would be no worse at estimating strangers' ages when those strangers have their hair disguised by a hat. Both experiments also examined the impact of disguising both features on age estimation bias but a lack of relevant past research precludes hypotheses.

Both experiments had secondary aims. Experiment 1 attempted to replicate earlier findings, showing young adults have an own-age advantage when estimating strangers' ages and that they overestimate the age of fellow young adults but underestimate the age of older adults (e.g., Voelkle et al., 2012). Experiment 2 attempted to replicate Dehon and Brédart's (2001) earlier finding that young to middle-aged White people are better at estimating the age of young to middle-aged White strangers than young to middle-aged Black strangers and that they overestimate the age of both. Experiment 1 and 2 also examined whether any own-age and own-race advantages that may be observed would stem from a lack of perceptual expertise in estimating the age of older adults and Black people. This was achieved by examining whether the amount of social contact participants had with older adults and Black people correlated with their accuracy when estimating those outgroup members' ages.

2 | EXPERIMENT 1

2.1 | Method

2.1.1 | Participants

There were 198 participants, aged 17–29 years (M_{age} = 20.48, SD = 2.88; F = 148, M = 50). Ninety-three had European ancestry, 63 East Asian ancestry, and 14 South Asian ancestry. The remainder had other ancestries (all n's < 6). All were studying psychology courses at a multi-campus university, were recruited via an online sign up system, consented to participate, and received course credit for participation. MorePower 6.04 (Campbell & Thompson, 2012) indicates this sample size has enough Power (>.80) to detect small interaction effects in the 4 × 4 repeated-measures ANOVA analysis used to examine age estimation accuracy (α = .05).

No age restrictions were enforced during recruitment and 30 additional people, aged 30–72, participated (M_{age} = 41.93; SD = 6.71; F = 16; M = 13; Unspecified = 1). They were excluded from all analyses as a secondary aim of Experiment 1 was to examine whether young adults are better at estimating the age of other young adults, relative to older adults. Here, an upper-age limit of 29 years was chosen for classifying someone as a young adult. Including the older participants in the analyses does not change the findings regarding the impact of disguises on age estimation accuracy.

2.1.2 | Design

Experiment 1 had a 4 × 4 repeated-measures design. The first independent variable was Disguise Type, with strangers wearing either no hat and sunglasses, a hat, sunglasses, or both. The second independent variable was age group, with participants estimating the ages of strangers in their twenties (i.e., fellow young adults), forties, sixties, or eighties. There were two primary dependent variables associated with age estimation. These were age estimation accuracy (measured by participants' MAE) and age estimation bias. Both were described earlier. A secondary dependent variable was the amount of social contact participants had with people in their twenties, forties, sixties, and eighties.

2.1.3 | Stimuli

Participants estimated the age of 32 White male strangers. The strangers appeared in individual passport-style colour photographs. Only their head and shoulders were visible. All had a neutral expression. None had facial adornments (e.g., glasses). All photographs were from databases created by DeBruine and Jones (2017) and Minear and Park (2004) for research purposes. Eight of the males were in their twenties (M_{age} = 22.50, *SD* = 1.60, Range 20–25), eight in their forties (M_{age} = 43.25, *SD* = 2.49, Range 40–47), eight in their sixties (M_{age} = 63.62, *SD* = 2.13, Range 61–66), and eight in their eighties (M_{age} = 81.87, *SD* = 1.72, Range 80–85). All photographs were digitally edited so the men appeared against a white background and were wearing either no sunglasses and a hat, a beanie-style hat, sunglasses, or the hat and sunglasses. Thus, there were four versions of each photograph (see Figure 1 for examples). The hat disguised the men's hair and forehead. All men had short hair, meaning none protruded from the back of their head/neck when "wearing" the hat but their sideburns were visible (as is the case when people wear a beanie). The sunglasses disguised their eye region, including their eyebrows. The sunglass lenses were black and opaque, so no information from the eye region was visible (as is the case with some styles, such as those worn by poker players).

Participants' social contact with people in their twenties, forties, sixties, and eighties was assessed via a Likert-scale type questionnaire adapted from Voci and Hewstone (2003). The original questionnaire includes five Likert-scale items assessing how much social contact people have with members of another race (e.g., A White participant may be asked "*How many Black people do you know very well?*", with the response options being "*Up to 2*", "*Up to 5*", "*Up to 8*", "*Up to 12*", *and* "*More than 12*"). The version used here substituted all references to race with references to the four age groups (e.g., "*How many people in their eighties do you know very well?*", with the response options being the same as above). The participants therefore answered the same five questions in relation to each age group, so answered 20 questions in total. The mean value of the responses to all five questions about an age group was used as a measure of social contact with that group, with higher scores indicating more contact. See the Supporting Information for the questions used.

2.1.4 | Procedure

Participants completed the study online and were asked to do so individually, at a quiet location of their choosing, on a laptop or personal computer. The study was hosted on Qualtrics. An information page initially informed participants the study was investigating their ability to estimate strangers' ages. There was no mention of the independent variables. After consenting to take part, participants completed a demographic questionnaire. They were then presented with each of the 32 strangers described above. Each stranger was presented individually and once only, in a fully randomised order, for 5 s. Participants always saw two strangers from each age group with no disguise, two with a hat, two with sunglasses, and two with a hat and sunglasses. The strangers in each disguise condition were counterbalanced across participants (e.g., Participant 1 saw a 20-year-old without a disguise, Participant 2 saw him with a hat only, Participant 3 saw him with sunglasses only, etc.). After studying a stranger for 5 s, participants were asked to estimate his age in years. A response had to be provided prior to seeing the next stranger. After estimating the age of all 32 strangers, participants completed the 20-item social contact questionnaire and the study ended. On average, the study lasted 13 min.

2.2 | Results

The data analysed in this experiment are available from the corresponding author.

2.2.1 | Age estimation accuracy

Age estimation accuracy was determined by calculating the MAE, or average age estimation error, within each of the 16 conditions (see Table 1). Across all conditions, the MAE was 7.68 years (*SD* = 5.27). Typically, a 4 × 4 repeated-measures ANOVA would be used to compare the MAE across all 16 conditions. Initial data screening, however, revealed the MAE data were positively skewed across all conditions, with Normal QQ-Plots suggesting it violated the normality assumption in all. The data were not, therefore, suited to parametric test analyses. Consequently,

FIGURE 1 An example of one of the strangers from Experiment 1 depicted without any disguise and the three types of disguise. The man pictured is forty

TABLE 1 Participants' mean absolute age estimation error in years when estimating the age of strangers from four age groups who were wearing either no disguise, a hat, sunglasses, or a hat and sunglasses

	Disguise type				
Age group	None	Hat	Sunglasses	Hat/sunglasses	*M* (*SD*)
Twenties	5.10 (3.44)	5.03 (3.15)	5.32 (3.26)	5.53 (3.50)	5.25 (3.34)
Forties	5.67 (3.19)	5.74 (3.48)	6.21 (3.01)	6.75 (4.24)	6.09 (3.53)
Sixties	7.90 (4.73)	8.18 (4.77)	8.69 (5.54)	9.24 (5.38)	8.50 (5.13)
Eighties	10.22 (6.60)	10.35 (6.73)	11.59 (5.96)	11.40 (6.47)	10.89 (6.45)
M (*SD*)	7.22 (5.11)	7.33 (5.18)	7.95 (5.21)	8.23 (5.51)	7.68 (5.27)

Note: SD's are in parentheses.

the data were transformed so a non-parametric 4×4 repeated-measures ANOVA could be used. More specifically, the data were transformed using an Aligned Rank Transformation via Wobbrock, Findlater, Gergle, and Higgins' (2011) ARTool (see Kay & Wobbrock, 2019, for an R package version). This tool aligns data (Hodges & Lehmann, 1962) and applies averaged ranks to it, meaning Factorial ANOVA procedures can be used to analyse it (Wobbrock et al., 2011).

The 4×4 non-parametric repeated-measures ANOVA revealed a statistically significant main effect of Disguise Type, $F(3, 591) = 14.05$, $p < .001$, $\omega^2 = .03$. Six multiple comparisons were then conducted to compare the MAE in all four disguise conditions. For all multiple comparisons in this article, a Holm correction (Holm, 1979) was applied to the observed *p* values to reduce the likelihood of Type 1 Errors. Adjusting the observed *p* values meant an alpha of .05 could be retained as the cut-off point for statistical significance (Chen, Feng, & Yi, 2017). MAE did not significantly differ when strangers wore no disguise or a hat only ($p = 1.00$). Similarly, it did not significantly differ when they wore sunglasses or sunglasses and a hat ($p = 1.00$). However, MAE in the latter two conditions (sunglasses; sunglasses and a hat) was significantly greater than in the former two (all four *p*'s < .001). In everyday terms, inaccuracy increased by just under a year when strangers wore sunglasses or sunglasses and a hat ($M = 8.09$ across both conditions), relative to when they wore no disguise ($M = 7.22$).

There was also a statistically significant main effect of Age Group, $F(3, 591) = 128.98$, $p < .001$, $\omega^2 = .31$. Multiple comparisons showed significant increases in MAE as the strangers increased in age from their twenties to forties, forties to sixties, and sixties to eighties (all *p*'s < .001). In everyday terms, inaccuracy doubled when strangers were in their eighties ($M = 10.89$), relative to their twenties ($M = 5.25$).

Finally, there was no statistically significant Disguise Type × Age Group interaction, $F(9, 1773) = 1.68$, $p = .09$, $\omega^2 = .003$. To summarise the main findings, age estimation accuracy decreased when the strangers were wearing sunglasses (irrespective of whether a hat was worn) and as the strangers' ages increased.

2.2.2 | Age estimation bias

The average age estimation bias in each condition can be seen in Figure 2. Across all conditions, participants underestimated strangers' ages

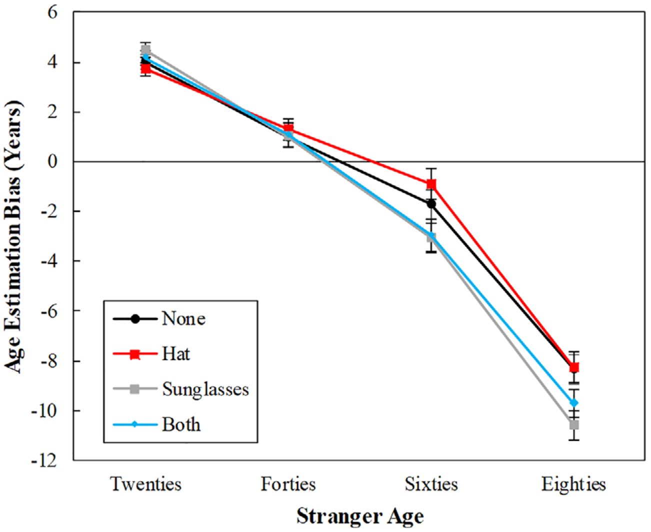

FIGURE 2 Participants' mean age estimation bias in years when estimating the age of strangers from four age groups when those strangers wore no disguise, a hat, sunglasses, or a hat and sunglasses. Error bars represent standard errors. 0 indicates no underestimation or overestimation

by an average of 1.55 years ($SD = 8.57$). Initial data screening revealed the age estimation bias scores had close to a normal distribution across all conditions but lacked sphericity (assessed via Mauchly's W). Consequently, a 4×4 repeated-measures ANOVA with a Greenhouse–Geisser correction was used to compare bias across conditions.

The ANOVA revealed statistically significant main effects of Disguise Type, $F(2.92, 574.41) = 6.59$, $p < .001$, $\omega^2 = .007$, and Age, $F(2.02, 398.56) = 394.13$, $p < .001$, $\omega^2 = .48$. These were qualified by a statistically significant Disguise Type × Age interaction, $F(6.87, 1{,}353.79) = 3.42$, $p = .001$, $\omega^2 = .007$. Among the myriad of multiple comparisons, it was found that participants overestimated the age of strangers who were in their twenties and there was a significant decrease in overestimation when strangers were in their forties, irrespective of Disguise Type (all *p*'s < .001). As the strangers increased in age from their forties to sixties, participants underestimated their age, regardless of Disguise Type, and the changes were significant (all *p*'s < .013). This underestimation was accentuated as strangers increased in age from their sixties to eighties and, regardless of Disguise Type, the changes were significant (all p's < .001). In addition, when strangers were in their twenties, age estimation bias did not significantly differ as a result of Disguise Type (all p's = 1.00). Similarly, when strangers were in the forties, age estimation bias did not significantly differ as a result of Disguise Type (all p's = 1.00). When strangers were in their sixties, underestimations were slightly greater if they wore sunglasses or sunglasses and a hat, relative to no disguise or just a hat. That pattern continued when strangers were in their eighties and underestimations were significantly greater if they wore sunglasses relative to a hat or no disguise (both p's < .001) but not if they wore sunglasses and a hat relative to just a hat or no disguise (both p's > .16, with both results being non-significant due to the Holm correction).

2.2.3 | Social contact

Secondary analyses focused on the amount of social contact participants had with people in their twenties, forties, sixties, and eighties. First, it was important to establish whether this differed for each age group. Initial data screening showed, as may be expected in a sample of young adults, that social contact scores were negatively skewed in relation to people in their twenties and forties (i.e., participants had high levels of social contact with people in these age groups) but positively skewed in relation to people in their sixties and eighties. Normal QQ-Plots suggested the data deviated from a normal distribution, so it was unsuited to parametric test analyses. A Friedman Test was, therefore, used to compare participants' social contact with each age group and it was found this significantly differed, $\chi^2_F(3) = 465.91$, $p < .001$, Kendall's $W = .43$. Multiple comparisons showed participants had significantly more contact with people in their twenties (*Mdn* = 4.80; IQR = .80) than forties (*Mdn* = 3.40, IQR = 1.40), forties than sixties (*Mdn* = 2.40, IQR = 1.20), and sixties than eighties (*Mdn* = 1.80, IQR = 1.20; all three p's < .001).

More importantly, three Spearman's rank correlations were conducted to examine if participants' social contact with people in their forties, sixties, and eighties correlated with their MAE when estimating their ages. To avoid potential confounds that may occur by collapsing MAE scores across the Disguise Type conditions, only participants' MAE in the no disguise conditions were examined. All correlations were small and non-significant (forties: $r = -.08$, $p = .28$; sixties: $r = .09$, $p = .23$; eighties: $r = .03$, $p = .65$).

2.3 | Discussion

As predicted, Experiment 1 found young adults' ability to accurately estimate strangers' ages declined if the strangers' eye region was disguised by sunglasses but not if their hair was disguised by a hat. This dovetails with Jones and Smith's (1984) earlier finding that children's accuracy when ranking strangers by age declined when the strangers' eye region, but not their hair, was disguised. Information from a stranger's eye region is, therefore, important for accurately estimating their age.

In addition, the young adult participants had an own-age advantage, with their estimations of strangers' ages becoming less accurate as the strangers' chronological ages increased. This own-age advantage replicates past findings (e.g., Anastasi & Rhodes, 2006; George & Hole, 1995; Klugman, 1947; Moyse & Brédart, 2012; Short et al., 2019; Thorley et al., 2018; Voelkle et al., 2012). The participants also overestimated the age of young adult strangers, had less bias when estimating the age of middle-aged strangers, and underestimated the age of the elderly strangers. This also replicates past findings (e.g., Henss, 1991; Short et al., 2019; Sörqvist & Eriksson, 2007; Vestlund et al., 2009; Voelkle et al., 2012; Willner & Rowe, 2001). A novel finding here is that the disguises had little impact upon this bias when strangers were in their twenties and forties. As the strangers' age increased further, however, the bias (i.e., underestimation) became progressively worse if they wore sunglasses or sunglasses and a hat.

Overall, the effect sizes showed that disguises had a small effect on age estimation accuracy, but a stranger's age had a large effect on age estimation accuracy. Age estimations, therefore, appear to be more influenced by a stranger's age than any of the disguises used here.

Finally, participants had less social contact with members of the older age groups but the amount of social contact they had with people from those groups and their accuracy when estimating the group members' ages did not correlate. Possible reasons for these null effects are considered in the General Discussion.

The primary aim of Experiment 2 was to replicate Experiment 1's finding that it is harder to accurately estimate a stranger's age if the stranger's eye region is disguised by sunglasses. A secondary aim was to replicate a past finding by Dehon and Brédart (2001) showing that young to middle-aged White people are better at estimating the age of young to middle-aged White strangers than young to middle-aged Black strangers. The White participants' social contact with Black people was also measured to see if greater levels of contact are associated with greater accuracy when estimating their ages.

3 | EXPERIMENT 2

3.1 | Method

3.1.1 | Participants

There were 112 participants, aged 17–37 years of age ($M_{age} = 20.97$, $SD = 4.58$; $F = 83$; $M = 29$). Only White people of European ancestry were eligible to participate in Experiment 2 and all confirmed their eligibility during testing. None took part in Experiment 1. All participants were studying psychology courses at a multi-campus university. All were recruited via an online sign up system, consented to participate, and received course credit for participation. MorePower 6.04 (Campbell & Thompson, 2012) indicates this sample size has enough Power (>.80) to detect a small-to-medium sized interaction effect in the 4 × 2 repeated-measures ANOVA analysis used to examine age estimation accuracy ($\alpha = .05$).

No age restrictions were enforced during recruitment and eight additional participants, aged 42–70, participated (M_{age} = 51.00; *SD* = 9.30; *F* = 3; *M* = 5). They were excluded from all analyses as they were older than the oldest of the strangers whose ages were being estimated in Experiment 2 (see below). This was done as one aim of Experiment 2 was to replicate Dehon and Brédart's (2001) earlier finding that young to middle-aged White people are more accurate when estimating the age of young to middle-aged White strangers than young to middle-aged Black strangers. In Dehon and Brédart's (2001) study, the participants were also no older than the strangers whose ages they were estimating. This was, therefore, kept consistent across both studies. Including the eight older participants in our analyses does not change the findings regarding the impact of disguises or race on age estimation accuracy.

3.1.2 | Design

Experiment 2 had a 4 × 2 repeated-measures design. The first independent variable was Disguise Type, with strangers wearing either none, a hat, sunglasses, or a hat and sunglasses. The second independent variable was Race, with participants estimating the age of Black strangers and White strangers. The two main dependent variables were age estimation accuracy and bias. A secondary dependent variable was the amount of social contact participants had with Black people and White people.

3.1.3 | Stimuli

Participants estimated the age of 24 male strangers. The strangers appeared in individual passport-style colour photographs taken from the same databases as Experiment 1. All had their head and shoulders showing, a neutral expression, and none had facial adornments. Twelve men were Black and 12 were White. The men from both racial groups were matched in age (M_{age} = 26.67, *SD* = 7.17, Range = 18–41). All images were edited as in Experiment 1 (e.g., the strangers were against a white background) and the same disguises applied again. This meant there were four versions of each photograph, with each stranger depicted wearing either no disguise, a hat only, sunglasses only, or a hat and sunglasses. All men had short hair that was fully covered when they “wore” the hat but their sideburns were visible.

Social contact with Black people and White people was assessed via the Likert-scale questionnaire described in Experiment 1 from Voci and Hewstone (2003). See the Supporting Information for the questions used.

3.1.4 | Procedure

Participants completed the study online and the procedures used mirrored those of Experiment 1. In brief, participants saw each of the 24 strangers, one at a time, in a fully randomised order, for 5 s. After 5 s, participants estimated the stranger's age in years. Participants always saw six strangers with no disguise, six with a hat, six with sunglasses, and six with a hat and sunglasses. Of the six strangers with each disguise type, three were always Black. The Black and White strangers with each disguise were counterbalanced across participants. After estimating the age of all 24 strangers, participants completed the social contact questionnaire. On average, the study lasted just over 9 min.

3.2 | Results

The data analysed in this experiment are available from the corresponding author.

3.2.1 | Age estimation accuracy

Age estimation accuracy was determined by calculating the MAE, or average age estimation error, within each of the eight conditions (see Table 2). The overall MAE was 6.33 years (*SD* = 3.00). The MAE data were positively skewed in all conditions and Normal QQ-Plots suggested the data violated the assumption of normality in most, but not all, of them. The data were not, therefore, suited to parametric test analyses. Consequently, the data were transformed using an Aligned Rank Transformation so a non-parametric 4 × 2 repeated-measures ANOVA could be used to compare the MAE across all conditions.

The non-parametric ANOVA revealed a statistically significant main effect of Disguise Type, $F(3, 333) = 7.59$, $p < .001$, $\omega^2 = .03$. Holm-corrected multiple comparisons showed MAE did not significantly differ when strangers wore no disguise or a hat only ($p = 1.00$). Similarly, it did not significantly differ when they wore sunglasses or sunglasses and a hat ($p = 1.00$). However, MAE in the latter two conditions (sunglasses; sunglasses and a hat) was significantly greater than in the former two (all four *p*'s < .005). In everyday terms, inaccuracy increased by just under a year when strangers wore sunglasses or sunglasses and a hat (*M* = 6.78 across both conditions), relative to when they wore no disguise (*M* = 5.87).

There was also a statistically significant main effect of Race, $F(1, 111) = 8.17$, $p = .005$, $\omega^2 = .01$, with MAE being greater when the White participants estimated the age of Black strangers. In everyday terms, their other-race age estimations were less accurate by just over half a year.

Finally, there was no statistically significant Disguise Type × Race interaction, $F(3, 333) = 0.04$, $p = .99$, $\omega^2 = <.001$. To summarise the findings, age estimation accuracy decreased when strangers wore sunglasses (irrespective of whether a hat was worn) and were from a different race.

3.2.2 | Age estimation bias

The average age estimation bias in each condition can be seen in Figure 3. Across all eight conditions, participants overestimated strangers' ages by average of 4.33 years (*SD* = 4.32). Initial data screening revealed the age estimation bias scores had close to a normal distribution across all conditions but lacked sphericity (assessed

TABLE 2 Participants' mean age estimation error in years when estimating the age of strangers from two racial groups who were wearing either no disguise, a hat, sunglasses, or a hat and sunglasses

	Disguise type				
Race	None	Hat	Sunglasses	Hat/sunglasses	*M* (*SD*)
Black	6.03 (2.95)	6.17 (2.78)	7.06 (3.23)	7.10 (3.48)	6.59 (3.15)
White	5.72 (3.15)	5.60 (2.46)	6.42 (2.82)	6.53 (2.81)	6.07 (2.84)
M (*SD*)	5.88 (3.05)	5.89 (2.63)	6.74 (3.06)	6.82 (3.17)	6.33 (3.00)

Note: SD's are in parentheses.

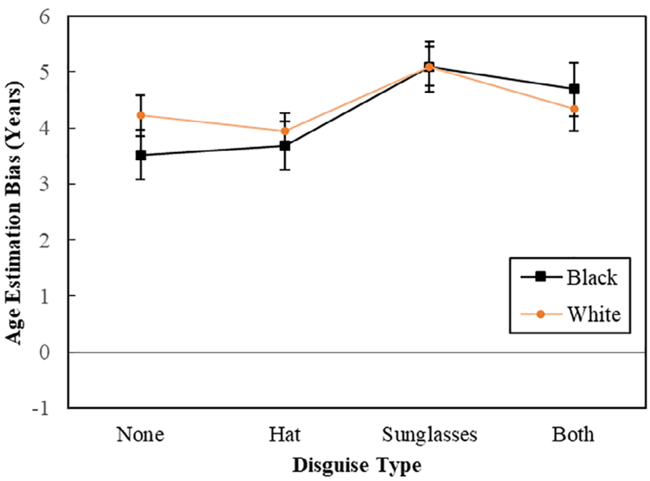

FIGURE 3 Participants' mean age estimation bias when estimating the age of Black strangers and White strangers when those strangers wore no disguise, a hat, sunglasses, or a hat and sunglasses. Error bars represent standard errors. 0 indicates no underestimation or overestimation

via Mauchly's W). Consequently, a 4 × 2 repeated-measures ANOVA with a Greenhouse–Geisser correction was used to compare bias across the conditions. There was a statistically significant main effect of Disguise Type, $F(2.77, 307.52) = 5.84$, $p < .001$, $\omega^2 = .02$. Holm-corrected multiple comparisons showed that when strangers wore no disguise or a hat, participants overestimated their ages and the degree of overestimation did not significantly differ ($p = .88$). Overestimation increased when strangers wore both a hat and sunglasses but not to the extent that it was significantly different from the aforementioned conditions (both *p*'s > .14). Overestimation increased further when strangers wore sunglasses only and this was significantly greater than when they wore no disguise or just a hat (both *p*'s < .004) but not when they wore both ($p = .31$). There was also no statistically significant main effect of Race, $F(1, 111) = 0.44$, $p = .51$, $\omega^2 = <.001$, or statistically significant Disguise Type × Race interaction, $F(2.79, 309.36) = 0.92$, $p = .42$, $\omega^2 = <.001$.

3.2.3 | Social contact

Secondary analyses focused on the amount of social contact White participants had with fellow White people and Black people. First, it is helpful to know if this differed. Initial data screening showed the mean social contact scores were negatively skewed in relation to White people but close to a normal distribution in relation to Black people, meaning they were not suited to parametric test analyses. A Wilcoxon Signed-Rank Test was therefore used to compare the two sets of social contact scores. It showed participants had significantly more social contact with White people (*Mdn* = 4.30, IQR = 1.00) than Black people (*Mdn* = 3.00, IQR = 1.20), $W = 6.50$, $p < .001$, $r = .10$.

More importantly, a Pearson's correlation examined if participants' social contact with Black people correlated with their ability to estimate Black strangers' ages. A Pearson's correlation was used as both variables had data with close to a normal distribution. To avoid a confound that may occur by collapsing MAE scores across Disguise Type conditions, only MAE scores in the no disguise condition were examined. The correlation was small and non-significant, $r = .13$, $p = .18$.

3.3 | Discussion

Experiment 2 replicated Experiment 1's finding that estimations of strangers' ages are less accurate when strangers have their eye region disguised by sunglasses but not when their hair is disguised by a hat. This again dovetails with Jones and Smith's (1984) earlier finding that children's ability to estimate strangers' ages declines when the strangers' eye region, but not their hair, is disguised and it confirms the importance of a stranger's eye region when estimating their age. Experiment 2 also replicated Dehon and Brédart's (2001) finding that young to middle-aged White adults are more accurate when estimating the age of young to middle-aged White strangers than young to middle-aged Black strangers, with participants overestimating the age of both. Here, it was also found that overestimations were greatest when the strangers wore sunglasses. When interpreting these findings, it is important to remember the effect sizes were small. Finally, White participants had less social contact with Black people and the amount of social contact they had with Black people did not correlate with their accuracy when estimating Black strangers' ages. This null effect is considered further in the General Discussion.

4 | GENERAL DISCUSSION

This research had three main findings. First, participants' estimations of strangers' ages were less accurate when the strangers' eye region was disguised by sunglasses but not when their hair was disguised by a hat. Second, the participants, who were mostly young White adults, had an own-age advantage (Experiment 1) and own-race advantage (Experiment 2) when estimating the strangers' ages, whereby their accuracy decreased as the strangers' chronological ages increased and when the strangers were Black. Third, the amount of social contact the participants had with older adults or Black people and the participants' accuracy when estimating their ages did not correlate. Each finding is discussed in turn next.

4.1 | Disguises and age estimation accuracy

Previous studies show estimations of non-disguised strangers' ages are often inaccurate by an average of close to 5 years (see Moyse, 2014, for a review). Here, comparable levels of inaccuracy were observed when young adult participants estimated the age of strangers who were not wearing any disguises and when those strangers were from their own age group. This occurred irrespective of the strangers' race.

Here, both experiments also showed that participants' accuracy when estimating strangers' ages declined when the strangers' eye region was disguised by sunglasses, and this occurred irrespective of the strangers' age and race. Moreover, it did not decline any further when the strangers' hair was also disguised by a hat. Effect size measures showed the decline in accuracy caused by the sunglasses was small. In everyday terms, across both experiments, the decline was by an average of just under 1 year. This decline in accuracy when sunglasses were worn was predicted as past research shows information from a stranger's eye region, such as the number of wrinkles around their eyes, is used to estimate their age (El Dib & Onsi, 2011; Lanitis, 2002; Merinville et al., 2015; Nkengne et al., 2008). Previously, however, only one study had demonstrated that it is harder to accurately estimate a stranger's age when their eye region is disguised. In that study, Jones and Smith (1984) found children's ability to rank photographs of child and adult strangers by age declined when the strangers' eye region was disguised. It is useful to know Jones and Smith's general observation replicates when very different samples, stimuli, and dependent measures are used.

Here, both experiments also found that participants' accuracy when estimating strangers' ages did not decline when the strangers' hair was disguised by a hat. Previous research had shown that information from a stranger's hair, such as the amount of grey present, can influence age estimations (Bulpitt et al., 2001; Butler et al., 1998; Gunn et al., 2009). Previous studies, however, had also shown that masking strangers' hair and head shape, irrespective of their gender, does not harm children and adults' ability to estimate the strangers' ages (George et al., 2000; George & Hole, 1995; Jones & Smith, 1984).

The above findings suggest that when a person is trying to estimate a stranger's age, information from the stranger's eye region is more helpful than their hair, given that disguising the former impairs age estimation accuracy but disguising the latter does not.

4.2 | Group characteristics and age estimation accuracy

Experiment 1 replicated earlier findings showing young adults are better at estimating the age of strangers from their own age group and that their accuracy declines as the strangers' chronological ages increase (Anastasi & Rhodes, 2006; George & Hole, 1995; Klugman, 1947; Moyse & Brédart, 2012; Short et al., 2019; Thorley et al., 2018; Voelkle et al., 2012). Here, effect size measures showed the effect of a stranger's age on accuracy was large and, in everyday terms, the young adult participants' inaccuracy doubled from an average of 5.25 years when strangers were in their twenties to an average of 10.89 years when strangers were in their eighties. Experiment 2 also replicated Dehon and Brédart's (2001) finding that young to middle-aged White people are better at estimating the age of young to middle-aged White strangers than young to middle-aged Black strangers. Effect size measures showed the effect of a stranger's race on age estimation accuracy was small and, in everyday terms, the decline in accuracy observed when the strangers were Black was, on average, close to half a year. A novel finding here is that these own-age and own-race advantages occurred regardless of whether the strangers wore disguises or not although, as mentioned, age estimation accuracy was generally worse when a disguise included sunglasses.

Researchers have considered why own-age and own-race age estimation advantages occur. One potential, but previously untested, explanation is that people have less social contact with members of other age groups and races, so lack perceptual expertise/proficiency in estimating their members' ages (Dehon & Brédart, 2001; George & Hole, 1995; Moyse & Brédart, 2012). If true, people who have regular social contact with other age groups and races should have developed perceptual expertise in estimating their members' ages and be better at it. In the present experiments, the (mostly) young White adult participants did have less social contact with older adults and Black people, but the degree of social contact they had with them did not correlate with their ability to accurately estimate these outgroup members' ages. Two possible reasons for these null effects are considered next.

First, there were only moderate amounts of variation in the amount of social contact the (mostly) young White adult participants had with people in their forties, sixties, and eighties and with Black people. For example, they generally reported having little contact with people in their eighties. It is therefore possible that this lack of variation can explain why social contact did not correlate with age estimation accuracy. Second, it is entirely possible that perceptual expertise cannot explain the own-age and own-race advantages observed here and in earlier studies. The perceptual-expertise explanation was inspired by findings from face recognition studies where own-age and own-race advantages in face recognition had been observed and were reduced/eliminated when participants had regular social contact with members of the outgroups whose faces they had to remember (e.g., Harrison & Hole, 2009; Meissner & Brigham, 2001; Rhodes & Anastasi, 2012). Importantly, manipulations that impair face recognition (such as inversion and photographic negation) leave age estimation accuracy relatively unaffected (George & Hole, 2000). It is therefore possible that the processes underpinning face recognition and age estimation differ and theoretical explanations for findings in one domain may not generalise to the other. It is also important to note that the own-age and own-race advantages, observed in the face recognition literature, are not solely due to differences in perceptual expertise, with social-cognitive factors also playing a role (see Young, Hugenberg, Bernstein, & Sacco, 2012). Whether these contribute to own-age and own-race age estimation advantages remain to be determined. It is beyond the scope of the current experiments to determine this but further direct tests of the perceptual-expertise explanation, alongside tests of other social-cognitive explanations, are encouraged.

4.3 | Age estimation bias

This research replicated past findings showing participants overestimate the age of young adults (irrespective of race), have less bias when estimating the age of middle-aged adults, and underestimate the age of elderly adults (e.g., Dehon & Brédart, 2001; Henss, 1991; Sörqvist & Eriksson, 2007; Vestlund et al., 2009; Voelkle et al., 2012; Willner & Rowe, 2001). The reason why this pattern of bias occurs is not well understood but see Voelkle et al. (2012) for a potential explanation centred on regression to the mean. In a novel finding, both experiments found disguises can impact upon age estimation bias. In Experiment 1, underestimation of the elderly strangers' ages increased when their eye region was disguised by sunglasses. A similar pattern was observed in Experiment 2, whereby overestimations generally increased when sunglasses were worn. Thus, disguising the eye region seems to accentuate age estimation bias in some conditions.

4.4 | Which factor mattered most?

The present experiments show the relative impact that a stranger's age, race, and their use of disguises can have on a person's ability to estimate their age. More specifically, the effect sizes show that when people attempt to estimate strangers' ages, the strangers' chronological age has the largest effect on their accuracy. The strangers' race and use of disguises, when those disguises are sunglasses and/or a hat, have a small effect on it.

4.5 | Limitations

The present experiments have several limitations that should be considered when interpreting their findings. One limitation is that participants estimated the age of strangers shown in passport-style photographs, and not 'live' strangers, so this reduces the experiments' ecological validity. When estimating the age of ‘live’ strangers there may be additional cues to their age, such as their clothing choices (Rexbye & Povlsen, 2007) and the sound of their voice (Moyse, 2014). However, Amilon et al. (2007) asked participants to estimate the age of strangers in a video where face and voice information were both available and the MAE was comparable to the no-disguise conditions in the current experiments. Similarly, Thorley et al. (2018) found sexual assault victims were able to estimate the age of stranger offenders during police interviews with an MAE comparable to the no-disguise conditions in the current experiments. It is therefore possible that the current findings may replicate if participants estimated the age of 'live' strangers who were wearing sunglasses and/or a hat.

Another limitation of the current experiments is that only one type of sunglasses and one type of hat were used as disguises. Here, the sunglasses disguised all information in the eye region, as is the case with some styles (e.g., those favoured by professional poker players). The hat covered the entire crown. Sunglasses, however, vary in how opaque they are and hats vary in the amount of hair they reveal. More revealing sunglasses and hats may result in more accurate age estimations.

A further limitation of the current experiments is that participants only estimated the age of male strangers. Research suggests people are better at estimating the age of male strangers than female strangers (e.g., Dehon & Brédart, 2001; Voelkle et al., 2012). It is therefore possible that the age estimation accuracy rates observed in the present experiments would be lower if participants estimated the age of female strangers.[1]

A final limitation of the current experiments is that the participants were mostly young White adults. This sample choice was deliberate as a secondary aim of these experiments was to replicate past findings showing an own-age advantage in young adults and an own-race advantage in young to middle-aged White adults. This does, however, mean that it is unknown whether the own-age and own-race advantages observed would be found in other age groups and races. The present experiments do not, therefore, offer evidence of own-age and own-race advantages in populations other than the current one, even if an own-age estimation bias has been found to occur across the lifespan (Anastasi & Rhodes, 2006; George & Hole, 1995; Klugman, 1947; Moyse & Brédart, 2012; Short et al., 2019; Thorley et al., 2018; Voelkle et al., 2012). It is feasible that older participants/non-White participants would have also been better at estimating the age of young White adult strangers (e.g., see Short et al., 2019, for evidence that young and elderly adult participants are sometimes better at estimating the age of young strangers, relative to elderly strangers).

4.6 | Applied implications

The present experiments' findings have potential implications for businesses and organisations that require people to accurately estimate strangers' ages. To illustrate this, two examples are offered.

First, in licensed bars, staff must decide whether customers who wish to purchase alcohol are legally old enough to do so. If an underage customer is sold alcohol in some countries, the staff and business can be fined (e.g., Australia, the UK). In the UK, the British Beer and

Pub Association is aware that underage people can appear old enough to purchase alcohol and bar managers have been asked to adopt the Challenge 21 policy. This policy requires staff to ask all customers who appear under 21 for identification proving they are old enough to purchase alcohol, even though the minimum age for purchasing alcohol is 18. This therefore provides a potential 3-year margin of error when estimating customers' ages. This policy has successfully reduced underaged alcohol sales (Home Office, 2007). Our findings, however, suggest staff could overestimate some underage customers' ages by more than 3 years, especially if the customers are from another race and wearing sunglasses (e.g., in outdoor bars). If so, this could result in underage sales. Adopting a wider margin of error may, therefore, be sensible (e.g., asking anyone who appears under 25 for identification). Research examining the effectiveness of different margins of error at preventing underage alcohol sales is, however, encouraged before policy recommendations are made.

Second, during police interviews, eyewitnesses who have seen a stranger offender commit a crime are often asked to estimate the offender's age[2] (Thomas, Aitken, Lucy, & Feist, 2004). The police then focus their investigation on potential suspects close to that age. If an eyewitness's age estimation is highly inaccurate, innocent people who are younger or older than the offender could be investigated, wasting police resources. Stranger offenders can, of course, come from any age group or race and some will deliberately try to disguise their appearance with sunglasses and a hat (Thorley et al., 2018; van Koppen & Lochun, 1997). The own-age estimation advantage observed here has been found to occur when eyewitnesses estimate the age of stranger offenders, meaning their estimations are less accurate as the strangers' chronological age increases (Thomas et al., 2004; Thorley et al., 2018). The present experiments suggest age estimation accuracy may decrease further if the eyewitness is White/the offender is Black and the offender was wearing sunglasses. If so, the police should bear this in mind when using age estimations to focus investigations on potential suspects close to an estimated age and alter their inclusion parameters accordingly. It is emphasised, however, that multiple studies replicating these effects, ideally with more ecologically valid stimuli, are recommended before any policy recommendations are made.

4.7 | Conclusion

The present experiments provided a much-needed investigation into how accurately people can estimate strangers' ages when the strangers' eyes are disguised by sunglasses and/or hair is disguised by a hat. Consistent with past research suggesting information from the eye region is used to estimate a stranger's age, it was found that age estimation accuracy decreased when strangers wore sunglasses but not when they wore a hat. The present experiments also replicated past findings showing people can have an own-age and own-race advantage when estimating strangers' ages. The experiments also provided an initial test of a previous suggestion that own-age and race advantages stem from a lack of social contact with other age groups and races. Importantly, there was no evidence in support of this suggestion.

CONFLICTS OF INTEREST

The author has no conflicts of interest.

ORCID

Craig Thorley https://orcid.org/0000-0002-3207-435X

ENDNOTES

[1] Readers may be interested to know there is little evidence that participant gender impacts upon age estimation accuracy (e.g., Dehon & Brédart, 2001; Voelkle et al., 2012). One study by Vestlund et al. (2009) is often reported as having found gender differences but a closer inspection of their results shows "*the main effect of Participant gender approached significance, F(1, 142) = 2.79, MSE = 1.56, p = .097,* η^2 *= 0.02*" (p. 305), but was not actually statistically significant. Nkengne et al. (2008) also reported that female participants' age estimation accuracy is slightly more precise than males, but they did not directly compare their male and female participants' accuracy. Here, at the request of an anonymous reviewer, an exploratory Mann–Whitney *U* test was ran comparing the male and female participants' age estimation accuracy when the male strangers wore no disguise. In Experiment 1, there were no statistically significant gender differences when participants estimated the age of strangers who were in their twenties, forties, sixties, or eighties (all p's > .05, all r_{rb}'s < .18). In Experiment 2, there were no statistically significant gender differences when participants estimated the age of White strangers or Black strangers (both p's > .014, all r_{rb}'s < .18). These findings therefore support earlier ones showing there are no gender differences in age estimation accuracy.

[2] While eyewitnesses estimate strangers' ages after a delay, and participants here did so while looking at strangers, evidence suggests age estimation accuracy may not degrade over time (e.g., Ebbesen & Rienick, 1998; Thorley et al., 2018; Yuille & Cutshall, 1986).

DATA AVAILABILITY STATEMENT

The data analysed in this experiment is available from the corresponding author.

ORCID

Craig Thorley https://orcid.org/0000-0002-3207-435X

REFERENCES

Amilon, K., Weijer, J. V. D., & Schötz, S. (2007). The impact of visual and auditory cues in age estimation. *Lecture Notes in Computer Science Speaker Classification II. 4441* 10–21. https://doi.org/10.1007/978-3-540-74122-0_2

Anastasi, J. S., & Rhodes, M. G. (2006). Evidence for an own-age bias in face recognition. *North American Journal of Psychology*, *8*(2), 237–252.

Aznar-Casanova, J., Torro-Alves, N., & Fukusima, S. (2010). How much older do you get when a wrinkle appears on your face? Modifying age estimates by number of wrinkles. *Aging, Neuropsychology, and Cognition*, *17*(4), 406–421. https://doi.org/10.1080/13825580903420153

Ebbesen E. B., & Rienick C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology*, *83*(5), 745–762. https://doi.org/10.1037/0021-9010.83.5.745

Bulpitt, C. J., Markowe, H. L. J., & Shipley, M. J. (2001). Why do some people look older than they should? *Postgraduate Medical Journal*, *77*(911), 578–581. https://doi.org/10.1136/pmj.77911.578

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society B: Biological Sciences*, *259*(1355), 137–143. https://doi.org/10.1098/rspb.1995.0021

Butler, J., Pryor, B., & Grieder, M. (1998). Impression formation as a function of male baldness. *Perceptual and Motor Skills*, *86*(1), 347–350. https://doi.org/10.2466/pms.1998.86.1.347

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Chen, S. Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease*, *9*(6), 1725–1729. https://doi.org/10.21037/jtd.2017.05.34

DeBruine, L., & Jones, B.. (2017, May 30). Face Research Lab London Set (Version 3). *figshare*. doi:https://doi.org/10.6084/m9.figshare.5047666.v3

Dehon, H., & Brédart, S. (2001). An 'other-race' effect in age estimation from faces. *Perception*, *30*(9), 1107–1113. https://doi.org/10.1068/p3122

El Dib, M. Y., & Onsi, H. M. (2011). Human age estimation framework using different facial parts. *Egyptian Informatics Journal*, *12*(1), 53–59. https://doi.org/10.1016/j.eij.2011.02.002

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior*, *27*(6), 433–442. https://doi.org/10.1016/j.evolhumbehav.2006.08.007

Fink, B., & Matts, P. J. (2008). The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology*, *22*(4), 493–498. https://doi.org/10.1111/j.1468-3083.2007.02512.x

George, P. A., & Hole, G. J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception*, *24*(9), 1059–1073. https://doi.org/10.1068/p241059

George, P. A., & Hole, G. J. (1998). The influence of feature-based information in the age processing of unfamiliar faces. *Perception*, *27*(3), 295–312. https://doi.org/10.1068/p270295

George, P. A., & Hole, G. J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition*, *7*(4), 485–509. https://doi.org/10.1080/135062800394621

George, P. A., Hole, G. J., & Scaife, M. (2000). Factors influencing young children's ability to discriminate unfamiliar faces by age. *International Journal of Behavioral Development*, *24*(4), 480–491. https://doi.org/10.1080/016502500750038035

Gunn, D. A., Rexbye, H., Griffiths, C. E., Murray, P. G., Fereday, A., Catt, S. D., …, Christensen, K. (2009). Why some women look young for their age. *PLoS ONE*, *4*(12), e8021. https://doi.org/10.1371/journal.pone.0008021

Harrison, V., & Hole, G. J. (2009). Evidence for a contact-based explanation of the own-age bias in face recognition. *Psychonomic Bulletin & Review*, *16*(2), 264–269. https://doi.org/10.3758/PBR.16.2.264

Henss, R. (1991). Perceiving age and attractiveness in facial photographs. *Journal of Applied Social Psychology*, *21*(11), 933–946. https://doi.org/10.1111/j.1559-1816.1991.tb00451.x

Hodges, J. L., & Lehmann, E. L. (1962). Rank methods for combination of independent experiments in analysis of variance. *The Annals of Mathematical Statistics*, *33*(2), 482–497. https://doi.org/10.1214/aoms/1177704575

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Home Office (2007, October). Underage Sales Down. https://web.archive.org/web/20080119233715/; http://press.homeoffice.gov.uk/press-releases/underage-sales-down

Jones, G., & Smith, P. K. (1984). The eyes have it: Young children's discrimination of age in masked and unmasked facial photographs. *Journal of Experimental Child Psychology*, *38*(2), 328–337. https://doi.org/10.1016/0022-0965(84)90129-2

Kay M., & Wobbrock. J. (2019). ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. R package version 0.10.6. doi:https://doi.org/10.5281/zenodo.594511.

Klugman, S. F. (1947). The effect of age upon estimation of age. *The Journal of Social Psychology*, *26*(1), 29–33. https://doi.org/10.1080/00224545.1947.9921729

Lanitis, A. (2002, July). *On the significance of different facial parts for automatic age estimation*. Paper presented at 2002 14th International Conference on Digital Signal Processing Proceedings, Volume 2 DSP 2002 (Cat. No. 02TH8628). IEEE. pp. 1027–1030. doi: https://doi.org/10.1109/ICDSP.2002.1028265

Mayes, A. E., Murray, P. G., Gunn, D. A., Tomlin, C. C., Catt, S. D., Wen, Y. B., … Granger, S. P. (2010). Ageing appearance in China: Biophysical profile of facial skin and its relationship to perceived age. *Journal of the European Academy of Dermatology and Venereology*, *24*(3), 341–348. https://doi.org/10.1111/j.1468-3083.2009.03418.x

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*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Merinville, E., Grennan, G. Z., Gillbro, J. M., Mathieu, J., & Mavon, A. (2015). Influence of facial skin ageing characteristics on the perceived age in a Russian female population. *International Journal of Cosmetic Science*, *37*(S1), 3–8. https://doi.org/10.1111/ics.12252

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633. https://doi.org/10.3758/BF03206543

Moyse, E. (2014). Age estimation from faces and voices: A review. *Psychologica Belgica*, *54*(3), 255–265. https://doi.org/10.5334/pb.aq

Moyse, E., & Brédart, S. (2012). An own-age bias in age estimation of faces. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *62*(1), 3–7. https://doi.org/10.1016/j.erap.2011.12.002

Nkengne, A., Bertin, C., Stamatas, G. N., Giron, A., Rossi, A., Issachar, N., & Fertil, B. (2008). Influence of facial skin attributes on the perceived age of Caucasian women. *Journal of the European Academy of Dermatology and Venereology*, *22*(8), 982–991. https://doi.org/10.1111/j.1468-3083.2008.02698.x

Porcheron, A., Latreille, J., Jdid, R., Tschachler, E., & Morizot, F. (2014). Influence of skin ageing features on Chinese women's perception of facial age and attractiveness. *International Journal of Cosmetic Science*, *36*(4), 312–320. https://doi.org/10.1111/ics.12128

Rexbye, H., Petersen, I., Iachina, M., Mortensen, J., McGue, M., Vaupel, J. W., & Christensen, K. (2005). Hair loss among elderly men: Etiology and impact on perceived age. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *60*(8), 1077–1082. https://doi.org/10.1093/gerona/60.8.1077

Rexbye, H., & Povlsen, J. (2007). Visual signs of ageing: What are we looking at? *International Journal of Ageing and Later Life*, *2*(1), 61–83. https://doi.org/10.3384/ijal.1652-8670.072161

Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology*, *23*(1), 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*(1), 146–174. https://doi.org/10.1037/a0025750

Samson, N., Fink, B., Matts, P. J., Dawes, N. C., & Weitz, S. (2010). Visible changes of female facial skin surface topography in relation to age and attractiveness perception. *Journal of Cosmetic Dermatology*, *9*(2), 79–88. https://doi.org/10.1111/j.1473-2165.2010.00489.x

Short, L. A., Mondloch, C. J., deJong, J., & Chan, H. (2019). Evidence for a young adult face bias in accuracy and consensus of age estimates.

British Journal of Psychology, *110*(4), 652–669. https://doi.org/10.1111/bjop.12370

Sörqvist, P., & Eriksson, M. (2007). Effects of training on age estimation. *Applied Cognitive Psychology*, *21*(1), 131–135. https://doi.org/10.1002/acp.1271

Thomas, N., Aitken, C., Lucy, D., & Feist, A. (2004). *Predicting the accuracy of stranger rape victims' statements* (Special interest series paper no 15). London: Home Office.

Thorley, C., Almond, L., Gregory, A., McAlonan, V., & McLoughlin, A. (2018). An archival analysis of sexual assault victims' age estimation accuracy when describing stranger offenders. *Psychology, Crime & Law*, *24*(10), 1030–1049. https://doi.org/10.1080/1068316X.2018.1503662

van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior*, *21*(6), 661–685. https://doi.org/10.1023/A:1024812831576

Vestlund, J., Langeborg, L., Sörqvist, P., & Eriksson, M. (2009). Experts on age estimation. *Scandinavian Journal of Psychology*, *50*(4), 301–307. https://doi.org/10.1111/j.1467-9450.2009.00726.x

Voci, A., & Hewstone, M. (2003). Intergroup contact and prejudice toward immigrants in Italy: The mediational role of anxiety and the moderational role of group salience. *Group Processes & Intergroup Relations*, *6*(1), 37–54. https://doi.org/10.1177/1368430203006001011

Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging*, *27*(2), 265–277. https://doi.org/10.1037/a0025065

Willner, P., & Rowe, G. (2001). Alcohol servers' estimates of young people's ages. *Drugs: Education, Prevention & Policy*, *8*(4), 375–383. https://doi.org/10.1080/09687630010019299

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011, May). *The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures*. Proceedings of the SIGCHI conference on human factors in computing systems, ACM. pp. 143–146.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116–142. https://doi.org/10.1177/1088868311418987

Yuille J. C., & Cutshall J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology*, *71*(2), 291–301. https://doi.org/10.1037/0021-9010.71.2.291.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Thorley C. How old was he? Disguises, age, and race impact upon age estimation accuracy. *Appl Cognit Psychol*. 2020;1–13. https://doi.org/10.1002/acp.3744