Checking ID-cards for the sale of restricted goods: age decisions bias face decisions

David J. Robertson[1, 2]*, & A. Mike Burton[1]

[1] Department of Psychology, University of York, York, UK

[2]School of Psychological Science & Health, University of Strathclyde, Glasgow, UK

Running Head

AGE AND IDENTITY VERIFICATION IN A RETAIL CONTEXT

***Corresponding Author**

Dr. David J. Robertson, School of Psychological Sciences and Health, University of Strathclyde, UK. E-Mail: david.j.robertson@strath.ac.uk Phone: +44 (0)141 548 4461

Conflict Of Interest Statement

The authors declare no competing interests

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.323262 to A. Mike Burton.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon request.

Abstract

Matching unfamiliar faces is highly error-prone, and most studies highlight the implications for real-world ID-checking. Here we study a particular instance of ID-checking: proof of age for buying restricted goods such as alcohol. In this case, checkers must establish that an identity document is carried by its legitimate owner (i.e. that the ID photo matches the face of the bearer) and that the ID proves the bearer to be old enough to make the purchase. Across three experiments, using two common forms of photo-ID (i.e. driving licences, PASS+ cards) we show that observers produce very high error rates when age requirements are met, but faces mismatch. This bias away from detecting a face mismatch remained evident in experienced cashiers – though to a somewhat attenuated level. We discuss interactions between face matching and other tasks, and the practical consequences of a bias which favours those using photo-ID with fraudulent intent.

Key Words

Identity Fraud, Face Recognition, Age Verification, Identity Cards, Expertise

Introduction

It is now well-established that matching pictures of unfamiliar faces is a challenging task and one which is highly prone to error (Bobak, Dowsett, & Bate, 2016; Bruce et al., 1999; Burton, White & McNeill, 2010; Fysh & Bindemann, 2018; Johnston & Edmonds, 2009, Robertson, Black, Chamberlain, Megreya, & Davis, 2020; Stacchi, Huguenin-Elie, Caldera, & Ramon, 2020). This difficulty is particularly evident when viewers are asked to match individuals to their photo-ID (Kemp, Towell & Pike, 1997; McCaffery & Burton, 2016; Meissner, Susa, & Ross, 2013; Papesh, 2018; Wirth & Carbon, 2017). Face matching is an important part of daily ID tasks, for example the purchase of age-restricted goods like alcohol and tobacco. However, the large majority of research on identification decisions focuses on the face match itself, not on other important biographical information present in ID documents. In the current study we examine age-verification from ID in a mock supermarket setting. Customers' ID must show a face that matches the bearer, and an age above the legal minimum for alcohol sales. We examine how these two task components combine and interact.

Our ability to recognise, or accurately match, new instances of people we are familiar with is an almost effortless and highly accurate process, even in cases in which there is considerable within person-variability across images (Jenkins, White, Van Montfort, & Burton, 2011). In contrast, when the individuals are unfamiliar to the observer, recognising or matching new instances of unfamiliar people is a challenging task and one in which error rates of 20% are common (Burton, White, & McNeill, 2010). Similarly high error rates are reported regardless of whether viewers are required to make a match/mismatch distinction between pairs of unfamiliar face photos (Megreya & Burton, 2006), or between a face photo and the live face of a person standing in front of them (Davis & Valentine, 2009; Megreya & Burton, 2008), as would be the case in real world contexts. Furthermore, substantial errors are also reported from those whose occupations involve face matching, such as police, passport officers or supermarket check-out staff (Burton, Wilson, Cowan, & Bruce, 1999; Kemp, Towell, & Pike, 1997; White, Kemp, Matheson, & Burton, 2014).

While our reliance on face photo ID for identity verification is not supported by psychological evidence, such documents remain the most widely used means of identity checking. Several studies have directly assessed unfamiliar face matching performance with face photos embedded in identity documents. Studies by Meissner, Susa, and Ross (2013) and Bindemann and Sandford (2011) have reported 20%-30% errors when viewers were asked to match face photos embedded in mock American passports, or in student ID cards, to a second face photo. In addition, Kemp, Towell and Pike (1997) reported large rates of error when asking supermarket cashiers to match face photos embedded in mock credit cards to the faces of customers live. While these studies further highlight the difficulties that unfamiliar face matching poses in applied contexts, they focused only on face matching. That is, the studies did not include experimental manipulations of the other biographical details on the identity document, such as the date or birth, which cashiers would use to verify the individual's age, or the biographical details, checked at passport control.

A recent study by McCaffery and Burton (2016) did directly investigate the interaction between unfamiliar face matching performance and the active assessment of related biographical information in a passport checking context. This study presented pairs of unfamiliar faces in isolation or with one of the faces embedded in a UK passport frame. Participants were asked to make match/mismatch decisions to the face pairs, but also to evaluate the biographical information when one of the faces appeared within a passport. Such data was sometimes implausible, for example noting the wrong sex for a person, or giving a date of birth which was highly implausible. Across three experiments, McCaffery and Burton (2016) reported that simply embedding a face in a passport frame consistently biased picture-matching responses towards 'match', i.e. participants missed more 'fraudulent' mismatch face pairs. Moreover, faces influenced data checking: when faces matched, participants were less likely to detect biographical errors. Further examination of face matching in real documents shows that the biasing effects of a passport context are also present in other documents such as driving licences and student ID (Feng & Burton, 2019).

These findings show that different sources of information in an identity document are not processed independently. This is important, because many uses of ID involve both face and biographical checks. For example, supermarket cashiers selling age-restricted goods need both to check identity and calculate age. Therefore, in this study we assess whether there are interactions between age and identity verification in a card-checking task. In our experimental context, participants take the role of a cashier who should ensure that only customers who are aged 18+ and who present a genuine ID document are sold alcohol (Experiment 1). We also assess identity checking accuracy (i.e. face matching performance) in this context using ID cards which explicitly represent age, rather than requiring an age calculation (Experiment 2). Finally, we compare the performance of student volunteers and individuals who currently work as supermarket cashiers on these tasks (Experiment 3). Across the three experiments, our focus is on error rates in our 'critical condition', which relates to ID checks in which an age appropriate card is presented but there is a mismatch between the face of the customer and the face photo on the identity card. This condition best mirrors ID fraud attacks in this context, in which minors are using someone else's identity card in order to try and obtain alcohol or other age restricted goods.

Experiment 1

In this experiment, participants are asked to take on the role of a cashier working at a supermarket checkout. They are informed that each of their customers would like to purchase alcohol, that the legal age for doing so is 18, and that in order to make the correct decision they must check the individual's ID card to verify their age and identity. Here we use mock-ups of UK provisional driving licences as the identity cards, as they can be obtained by individuals as young as 15 years 9 months, and are a widely-accepted form of ID. These ID cards display dates of birth, which we manipulated here to denote ages between 16 and 26 at the time of the study. Within the experiment, the participants would encounter customers whose faces matched or did not match the photo-ID, and ID cards showing the bearer to be old enough or too young to buy alcohol. A common real-world fraud is mimicked by the condition in which an ID denotes legal age but the face mismatches. A minor could obtain alcohol by presenting another person's ID card showing an individual who looks somewhat like them, and which verifies that they are old enough to buy alcohol.

Method

Ethics Statement

Each experiment reported in this paper was approved by the Ethics Committee of the Department of Psychology, University of York, UK. All participants provided written informed consent, had normal or corrected-to-normal vision, and each received a course credit or monetary payment for their participation.

Participants

Thirty-six participants (30 female) with a mean age of 21 years (*SD* = 3, *Range* = 18-31) were recruited from the University of York Department of Psychology. None of the participants currently worked full-time or part-time in an occupation that required photo-ID identity checks.

Stimuli and Apparatus

One hundred and forty-four pairs of faces were selected from the Glasgow Face Matching Test (GFMT; Burton, White & McNeill, 2010), half of the pairs showed two photos of the same person (face match trials) and the remaining half showed two different photos of similar looking people (face mismatch trials). Faces in the GFMT show young people who were students at the time of photography. Two images were created for each of the faces in the set. The first image was a large (7cm x 9cm) full colour photo, which provided the images of the 'customer'. The second image was a smaller black and white photo which was embedded in a mock UK Provisional Driving Licence frame, using the same dimensions as a real licence (frame 8.5cm x 5.5cm; photo 3.5cm x 4.5cm). As seen in Figure 1, the driving licences used in the experiment closely matched their real-world counterparts. On each trial, the customer's face photo and their identity card were displayed within a supermarket cashier context, with an image showing shelves of alcohol in the background, and the checkout setup in the foreground. The experiment was presented on a 12-inch Hewlett Packard laptop using E-Prime 2.0.

Procedure

The experiment began by outlining the context of the task. Participants were told that they would take on the role of a supermarket cashier working at the checkout. The participants were told that each of the customers they would face would be trying to buy alcohol, and on each occasion (i.e. trial), the customer would present the cashier with an ID card which included a face photo and a date of birth. Participants were reminded that the sale of alcohol to anyone under the age of 18 was illegal, and that they should only agree to the sale if the ID card verified the customer's identity (i.e. the face photo on the card matched the customer's face) and their age (i.e. their date of birth made them aged 18 or above). Participants were made aware of the location of the face photo and the date of birth in an example driving licence presented at the start of the task.

Trials were self-paced, and participants made a keyboard button-press to indicate whether they would, or would not, sell alcohol in this case. Across the task, participants encountered all combinations of face (match/mismatch) and age (old enough/too young). Each testing session consisted of 144 trials, with 36 trials per experimental condition. For the 72 trials in which the ID card showed the owner to be too young, a date of birth was randomly selected which indicated that the bearer was 16 or 17. For the remaining 72, dates of birth in the age range of 18-26 years were randomly selected. Trial order was randomised within participants, while face image was counterbalanced across participants such that both identities within a face pair were shown both as the 'customer's' face and as the face photo on the identity card. In addition, the irrelevant information that appeared on the ID card also varied trial to trial, with name, driver number, city, address, post code, date of issue and date of expiry all being randomly selected in order to make the context as realistic as possible.

Across the three experiments reported in this paper, participants were also asked to record their ID checking strategy on completion of the behavioural task. The response options were 'I checked the face first followed by the date of birth', or 'I checked the date of birth first followed by the face'. A single analysis is performed on this data across the three experiments and is reported at the end of the results section for Experiment 3. A typical testing session lasted approximately 45 minutes.

--- FIGURE 1 HERE PLEASE ---

Results and Discussion

Response Errors (%)

Figure 2 shows participants' mean response errors across condition. These were entered into a 2 × 2 repeated measures ANOVA with the factors of matching condition (face match, face mismatch) and age denoted on card (under 18, over 18). The ANOVA revealed main effects of both the matching, $F(1, 35) = 54.51$, $p < .001$, $\eta_p^2 = .61$, and the age conditions, $F(1, 35) = 74.95$, $p < .001$, $\eta_p^2 = .68$, which were qualified by a matching × age interaction, $F(1, 35) = 177.63$, $p < .001$, $\eta_p^2 = .84$. Tests of simple main effects showed that for the under 18 cards, response errors in the face match condition were significantly higher than in the face mismatch condition, $F(1, 35) = 9.94$, $p = .003$, $\eta_p^2 = .21$. In contrast, an opposite and much larger effect was observed for over 18 cards, in which response errors in the face mismatch condition were significantly higher, by a margin of 46% on average, compared to the face match condition, $F(1, 35) = 105.04$, $p < .001$, $\eta_p^2 = .75$.

The most striking effect here is that participants shown a legal-age driving licence, are highly prone to making an error when the faces mismatch. The decision to sell alcohol seems to be heavily biased towards use of the age data, and not evidence from face matching – a conclusion also consistent with the generally low error rate for under-18 cards. In short, use of a fraudulent ID card, showing someone of legal age, is hard for our participants to detect.

--- FIGURE 2 HERE PLEASE ---

Response Times (RTs)

A second ANOVA on participants' mean correct response times (RTs) revealed a main effect of age, $F(1, 35) = 11.30$, $p = .002$, $\eta_p^2 = .24$, but not matching condition, $F(1, 35) = 2.99$, $p = .093$, $\eta_p^2 = .79$, which was qualified by an age × matching condition interaction, $F(1, 35) = 13.46$, $p = .001$, $\eta_p^2 = .28$. Follow up tests of simple main effects revealed the same pattern as reported for response errors above, with significantly longer RTs for under 18 card condition when there was a face match ($M = 3.0$s), compared to a face mismatch ($M = 3.4$s), $F(1, 35) = 6.43$, $p = .016$, $\eta_p^2 = .16$. In contrast, for the over 18 card condition, RTs were significantly longer in the presence of a face mismatch ($M = 4.6$s), compared to a face match ($M = 3.0$s), $F(1, 35) = 8.25$, $p = .007$, $\eta_p^2 = .19$.

Experiment 2

In Experiment 1, we used a typical UK identity card, the provisional driving licence, and the results showed that the relatively simple process of an age calculation appears to bias subsequent identity verification decisions, leading to more acceptances of faces that do not match. This is a problem for security, as it suggests that people with fraudulent intent could benefit from use of a card showing legitimate age, even if the face is not a good match. One reason for the poor performance reported in Experiment 1 could be the level of cognitive load imposed on checkers as a result of having to perform a numerical calculation and a demanding unfamiliar face matching task. Limited capacity processing has been shown across a wide variety of tasks (Lavie, 1995; Lavie, Hirst, de Fockert, & Viding, 2004), and it could be the case that working memory capacity limits are the underlying cause for such poor decision making (Baddeley, 1992; Anderson, Reder, & Lebiere, 1996). Therefore, in Experiment 2 we seek to assess whether the removal of the age calculation, and the associated reduction in cognitive load imposed by the checking process, could ameliorate this effect.

To that end, in Experiment 2, we test ID cards which explicitly state that the bearer is of legal age, rather than requiring an age calculation from d.o.b. PASS+ is a 'Proof of Age Standards Scheme' set up in the UK in 2014 (http://www.pass-scheme.org.uk/). Identity cards are issued to individuals who have proved during a rigorous application process that they are indeed aged 18 or above. This application process involves, among other verification steps, the submission of official documents (e.g. birth certificate, passport, NHS card, CRB check) countersigned by an appropriate referee (see https://www.citizencard.com/requirements-for-a-first-uk-id-card). The cards bear the owner's photograph (similar to a driving licence) and confirm that the owner is over 18 (see Figure 3). So, people checking legal age for purchase of restricted goods need only ensure that the photo matches the bearer. In Experiment 2 we examine identity checking errors in the same sale of alcohol context used in Experiment 1, comparing ID cards which require an age calculation (UK Provisional Driving Licence) and those which do not (PASS+ Cards).

Method

Participants

Thirty-six participants (29 female) with a mean age of 20 years (*SD* = 3, *Range* = 17-32) were recruited from the University of York Department of Psychology. None of the participants currently worked full-time or part-time in an occupation that required photo-ID identity checks.

Stimuli, Apparatus and Procedure

The stimuli, apparatus and procedure were identical to those described for Experiment 1, with the exception that half of the customers (72 trials, 36 Face Match, 36 Face Mismatch) now showed a PASS+ identity card. Therefore, the initial instructions were amended to inform the participants that both age and identity had to be verified when the customer presented a provisional driving licence, but only identity had to be verified if the customer presented an 18+ PASS card.

--- FIGURE 3 HERE PLEASE ---

Results and Discussion

Response Errors (%)

Figure 4 shows participants' mean response errors across condition. The driving licence condition provided a direct replication of our design from Experiment 1, so here we test that replication using a 2 x 2 repeated measures ANOVA on mean response errors, with the factors of face matching condition (match, mismatch) and age on card (under 18, over 18). The ANOVA revealed main effects of both the matching, $F(1, 35) = 50.76$, $p < .001$, $\eta_p^2 =$

.59, and age conditions, $F(1, 35) = 103.70$, $p < .001$, $\eta_p^2 = .75$, which were qualified by a matching × age interaction, $F(1, 35) = 75.81$, $p < .001$, $\eta_p^2 = .68$. Follow up tests of simple main effects showed an identical pattern of findings to that reported in Experiment 1. For the under 18 cards, response errors in the face match condition were significantly higher than in the face mismatch condition, $F(1, 35) = 13.87$, $p = .001$, $\eta_p^2 = .28$. In contrast, an opposite and much larger effect was observed for over 18 cards, in which response errors in the face mismatch condition were significantly higher, by a margin of 38% on average in this Experiment, compared to the face match condition, $F(1, 35) = 69.84$, $p < .001$, $\eta_p^2 = .67$. Note that the typical GFMT mismatch condition error rate is ~10%.

Following the replication of our results from Experiment 1, we next examined performance on the PASS+ card, which confirms legitimate age for alcohol purchase. As can be seen in Figure 4, the PASS+ card gives very similar results to that of the legitimate-age Driving Licence: error rates were significantly higher in the face mismatch condition compared to the face match condition, $t(35) = 9.63$, $p < .001$, $d = 1.67$. That is, the use of the PASS+ card did not eliminate the age verification bias. As seen in Figure 4, for both match, $t(35) = 5.28$, $p < .001$, $d = .88$, and mismatch conditions, $t(35) = 2.42$, $p = .021$, $d = .40$, error rates were significantly lower for PASS+ cards compared to driving licences. However, these reductions were small in size, with only a 7% reduction for PASS+ cards compared to driving licences in the critical condition (18+, faces mismatch). That is, when the requirement to make an age calculation is removed, and the cognitive load of the task is reduced, there is an 7% reduction in the likelihood that a cashier would accept an ID card in which the face photo does not match the customer's faces. However, the reduction in error is modest in size, and

38% errors in the critical condition in the PASS+ card still represents a threefold increase in error compared to the typical rates generated by the regular GFMT (i.e. face matching with no ‘ID card’ context). This may suggest, that the mere presence of a face in an official identity document used to explicitly verify the bearer’s age, could be enough to significantly bias decisions in favour of fraudulent use (see McCaffery & Burton, 2016).

--- FIGURE 4 HERE PLEASE ---

Response Times (RTs)

A second 2 × 2 repeated measures ANOVA on mean correct response times for the driving licence conditions revealed no main effects and no interaction (all *F’s* < 1). RTs across the conditions were relatively consistent: for the under 18 cards, mean correct RTs were 4.5s in the match condition and 4.2s in the mismatch condition. For over 18 cards, mean correct RTs were 4.4s in the match condition and 4.6s in the mismatch condition. While there were no significant effects on RTs in Experiment 2, the data does trend in the same direction as reported in Experiment 1. In addition, RTs across these conditions were at least a full second longer than those reported for the equivalent conditions in Experiment 1. It could be the case that due to the inclusion of additional face matching content in the instructions of the present experiment, that greater attention was paid to the matching element of the task. Despite this, mean response error rates do not differ substantially between Experiment 1 and

Experiment 2, suggesting that it is an age verification bias rather than a speed/accuracy trade-off that is driving the effect on error rates.

As with the response error analysis, the important comparisons here were within the PASS+ card conditions (match, mismatch), and between the critical conditions (18+, mismatch) across card type. Paired t-tests revealed that, for the PASS+ card, correct RTs were significantly longer in the face mismatch condition (*M* = 4.7s), compared to the face match condition (*M* = 3.3s), supporting the findings from Experiment 1, for driving licences, in which an over 18 card with mismatching faces required the greatest deliberation time and cognitive effort. As with the error rate findings, eliminating the age calculation through the use of the PASS+ card, did not eliminate this effect. Across card types, correct RTs were found to be significantly faster in the PASS+ card match condition (*M* = 3.3s), , compared to the driving licence match condition (*M* = 4.4s), $t(35) = 3.68$, $p = .001$, $d = .62$. Taken together with the error rate analysis, this shows that using an PASS+ card leads observers to make the correct decision more quickly and more often when the bearer's face is a match to the face on the card. However, somewhat surprisingly, there was no difference in correct RTs between the PASS+ card mismatch condition (*M* = 4.7s) and the 18+ mismatch driving licence condition (*M* = 4.6s), $t < 1$. This suggests that the age confirmation bias remains, in relation to RTs, regardless of whether an active calculation is required or not.

Experiment 3

Experiment 2 replicated our original study in showing that legal-purchase checks are severely compromised when ID shows legitimate age. In both Experiments 1 and 2, people are highly likely to accept ID as genuine, even with a mismatching face, as long as the bearer's age is legitimate. Experiment 2 shows that this effect is not due simply to ID-checker's having to calculate an age from a d.o.b. Even when age-legitimacy is given in the PASS+ card, leaving a requirement only to compare faces, participants make very high numbers of errors. Note that the face pairs used here come from a standardised face test, the Glasgow Face Matching Test. Errors on this test, when showing face pairs in isolation (i.e. without ID cards or backgrounds) are typically in the range of 10-18% (Burton, White, & McNeill, 2010; Fysh & Bindemann, 2018; Robertson, Black, Chamberlain, Megreya, & Davis, 2020; McCaffery, Robertson, Young, & Burton, 2018; Verhallen et al., 2017). So, in these experiments' errors are much higher, even in conditions where context is essentially task-irrelevant, because viewers have only to match faces.

While these experiments suggest that photo-ID does not guarantee compliance with age-restricted purchases, we do not yet know whether this would be a problem for experienced check-out staff. Several studies have shown that occupational experience does not appear to be associated with identity verification performance, with police officers (Burton, Wilson, Cowan, & Bruce, 1999; Wirth & Carbon, 2017), and passport officers (White, Kemp, Jenkins, Matheson, & Burton, 2014) performing no better than untrained student controls. Further, Kemp et al. (1997) showed high levels of face matching error, without an age calculation, in supermarket cashiers asked to verify face-photo credit cards. However, we need to establish whether experience is a key issue for the specific task under study here – checking age-appropriate purchases using standard photo-ID. In the following experiment, we recruited experienced supermarket check-out staff. We also returned to the use of driving licence-ID only, as this is by far the most common form of ID-verification at purchase points.

Method

Participants

Eighteen student cashiers (14 Female) with a mean age of 20 years (*SD* = 1, *Range* = 19-24) were recruited from a University of York research advertising service. From this sample, all cashiers were currently active in part-time jobs requiring age and identity verification to ensure that alcohol was not sold to minors. In terms of occupation, 11 cashiers reported working in a supermarket/shop, and 7 in a bar/pub. The mean number of months in which the cashiers had been employed in these roles was 15 (SD = 12), 2 cashiers had only been in post for 1 month but confirmed that during that period they had requested ID from a customer on at least one occasion. A further eighteen participants (16 Female) with a mean age of 19 years (*SD* = 1, *Range* = 18-20) were recruited from the University of York participant pool, these individuals had no prior experience an any role that required identity checking. Our sample was well matched for sex, and while the cashier group was statistically older than the control group, $t(34) = 2.73$, $p = .010$, $d = .94$; numerically, this was a modest difference of one year on average.

Stimuli, Apparatus and Procedure

The stimuli, apparatus and procedure were identical to those reported in Experiment 1, with the exception that the cashier group provided additional information on the type of ID checking role in which they worked, and the amount of time they had spent doing so.

Results and Discussion

Response Errors (%)

Figure 5 shows participants' mean response errors across condition. Data were entered into a 2 × 2 × 2 mixed design ANOVA with the within subjects factors of matching condition (face matching, face mismatch) and age on card (over 18, under 18), and the between subjects factor of group (cashiers, controls). The ANOVA revealed that each of the main effects and two-way interactions were significant, full statistics for these effects are included in the supplementary materials. These effects were qualified by an matching × age × group interaction, $F(1, 34) = 6.07$, $p = .019$, $\eta_p^2 = .15$. As a result of this three-way interaction, and for clarity in our reporting of the results, we will split the data by group and perform two 2 x 2 repeated measures ANOVAs with the factors of with the factors of matching condition (face match, face mismatch) and age denoted on card (under 18, over 18).

For the control group, the ANOVA revealed main effects of both the matching, $F(1, 17) = 56.59$, $p < .001$, $\eta_p^2 = .77$, and the age conditions, $F(1, 17) = 72.69$, $p < .001$, $\eta_p^2 = .81$, which were qualified by a matching × age interaction, $F(1, 17) = 97.29$, $p < .001$, $\eta_p^2 = .85$. As seen in Figure 5, tests of simple main effects showed that for the under 18 cards, there was a trend towards response errors in the face match condition being higher than in the face mismatch condition, $F(1, 17) = 3.14$, $p = .094$, $\eta_p^2 = .16$. In contrast, an opposite, larger, and significant effect was observed for over 18 cards, in which response errors in the face mismatch condition were significantly higher, by a margin of 51% on average, compared to the face match condition, $F(1, 17) = 75.59$, $p < .001$, $\eta_p^2 = .82$. These results from the control group, replicate our findings for over 18 driving licence cards reported in Experiment's 1 and 2. The non-significant difference between match and mismatch for under 18 cards in this Experiment is likely to be due to the reduced sample size ($N = 18$) in comparison to Experiment's 1 and 2 (both $N = 36$).

For the cashier group, the ANOVA revealed main effects of both the matching, $F(1, 17) = 14.84$, $p = .001$, $\eta_p^2 = .77$, and the age conditions, $F(1, 17) = 15.64$, $p = .001$, $\eta_p^2 = .48$, which were qualified by a matching × age interaction, $F(1, 17) = 39.03$, $p < .001$, $\eta_p^2 = .70$. As seen in Figure 5, and in line with the effects reported for the control group, tests of simple main effects showed that for the under 18 cards, there was a non-significant trend, $F < 1$, towards response errors in the face match condition being higher than in the face mismatch condition. In contrast, an opposite, larger, and significant effect was observed for over 18 cards, in which response errors in the face mismatch condition were significantly higher, by a margin of 30% on average, compared to the face match condition, $F(1, 17) = 45.01$, $p < .001$, $\eta_p^2 = .73$. These results show that the same bias found in the control group and in the untrained samples recruited for Experiment's 1 and 2 were still present in the cashier group. That is, experienced and trained cashiers who are employed in roles which require age and identity checks were subject to the same effects found in an untrained sample of university students.

Having established that the same bias exists in both cashiers and controls, it is important to assess whether the reduced cashier error rates represent a significant effect. A series of independent t-tests showed that while there were no differences between the groups for match and mismatch error rates in the under 18 card condition, *t's* < 1, and no difference between groups in the match condition for over 18 cards, *t* < 1, there was a significant difference in performance between cashiers and controls in the critical over 18 card mismatch condition, $t(34) = 3.06$, $p = .004$, $d = 1.02$, with the mean cashiers error rate being 21% lower than the control group, as seen in Figure 5. That is, in a fraud situation in which a customer has presented a driving license to a cashier with a face photo of another person, with a d.o.b which would make them over 18, cashiers would wrongly sell them alcohol 34% of the time compared to 55% in controls. Although there is a cashier advantage here, error rates of 34% in this context are still unacceptable high, and these mismatch error rates are much higher than one would expect if the faces were presented in isolation.

Response Times (RTs)

A second 2 × 2 × 2 mixed design ANOVA on participants mean correct response times (RTs) revealed no main effect of group, $F < 1$, no interaction between group and face matching or age on card, both *F's* < 1, and no three-way interaction between group, face matching and age, $F(1, 34) = 1.16$, $p = .289$, $\eta_p^2 = .03$. However, there was a main effect of age on card, $F(1, 34) = 19.93$, $p < .001$, $\eta_p^2 = .37$, and face matching condition, $F(1, 34) = 15.25$, $p < .001$, $\eta_p^2 = .31$, which were qualified by an age × matching condition interaction,

$F(1, 34) = 17.71$, $p < .001$, $\eta_p^2 = .34$. Follow up tests of simple main effects revealed the same pattern as reported for response errors above, with no significant difference in RTs for the under 18 card condition when there was a face match ($M = 3.5$s), compared to a face mismatch ($M = 3.4$s), $F < 1$. In contrast, for the over 18 card condition, RTs were significantly longer in the presence of a face mismatch ($M = 5.3$s), compared to a face match ($M = 3.5$s), $F(1, 35) = 21.6$, $p < .001$, $\eta_p^2 = .38$.

--- FIGURE 5 HERE PLEASE ---

Cashier Experience

A Pearson's correlation analysis on the number of months cashiers had spent in an ID checking role and their critical condition error rates revealed a non-significant positive association, $r(18) = .412$, $p = .089$. There is no hint here of an expertise effect in which more experienced cashiers perform better – with the (non-significant) trend going in the opposite direction.

This study shows reduced error rates by experienced cashiers (as compared to inexperienced students) in the critical 'correct-age/wrong photo' condition. However, the cashiers' error rates remain very high. Once again, we note that face matching with these stimuli (the GFMT) typically give rise to error rates well below 20%, but that embedding the decision within an age-check purchase context gives rise to much higher rates – even for those experienced with the task.

ID Checking Strategy

As noted above, across the three experiments, participants were asked to record their ID checking strategy for the driving licence ID cards upon completion of the checking task. The response options open to participants were ‘I checked the face first followed by the date of birth’, or ‘I checked the date of birth first followed by the face’. Here we collapse strategy responses from participants across the three experiments and assess whether checking strategy preference affects critical condition error rates. Across the sample (*N* = 108), 75 participants (69%) reported adopting an age-then-face ID checking strategy, with the remaining 33 participants (31%) opting to check the face first followed by the age. There was no difference in the number of participants who adopted either strategy between the cashier and control groups in Experiment 3.

An independent-samples t-test revealed that significantly fewer critical condition errors were made by those participants who adopted a Face-Age checking strategy (*M* = 26%, *SD* = 18%, *Range* = 0%-66%), compared to an Age-Face checking strategy (*M* = 52%, *SD* = 22%, *Range* = 6%-100%), $t(106) = 6.08$, $p < .001$, $d = 1.29$. While it is the case that the checking strategy groups have unequal sample sizes, this finding provides preliminary evidence for the view that the adoption of a face first ID checking strategy, may be advantageous in reducing the likelihood that a cashier would accept a mismatching, but age appropriate, identity card.

General Discussion

This was the first study, to our knowledge, to assess interactions between age and identity verification in a retail context, using novice undergraduate students and experienced cashiers. Across three experiments we show that successfully verifying age, on the basis of standard personal-ID, significantly biased observer judgements away from spotting face mismatches. Here, we show that, for untrained observers, 51% (Experiment 1), 45% (Experiment 2) and 45% (Experiment 3) of the time that a participant was asked to detect a 'fraud' (i.e. critical condition 18+/face mismatch trials), they made the wrong decision, leading to the illegal say of alcohol. Typical error rates with these facial stimuli range between 11% and 18%, depending on which subset is used (Burton et al, 2010). So, these error rates are much higher than one would expect based on performance with the faces in isolation.

This finding is line with the results reported by McCaffery and Burton (2016). They reported, in a passport checking context, that a 'same' face matching judgement resulted in the detection of fewer errors in related biographical details. Similarly, here we show that an accurate 18+ age verification calculation resulted in the detection significantly fewer mismatch face pairs. Across both studies, these interactions between face matching and other related information biases the final checking decision in an unsafe direction (i.e. favouring users of fraudulent documents). Here we characterise this effect as an age confirmation bias (Klayman, 1995; Mynatt, Doherty, & Tweney, 1977; Nickerson, 1998), in which the cognitive resources that have gone into accurately calculating an 18+ age from date of birth information, creates a pre-existing framework (e.g. 'this ID check is likely to be fine') which biases the face matching decision to accept that the customers face matches the face photo, when in fact they are two different people.

In Experiment 2, we found that eliminating the cognitive demands of an age calculation, through the use of pre-verified PASS+ cards, reduced, but did not eliminate critical condition errors, compared to the driving licence ID's. This finding showed that the age conformation bias appears to persist, even when the cognitive load imposed by the checking task is significantly reduced. In other words, the pre-existing age verification framework, this time imposed by the nature of the PASS+ card and not via a cognitively demanding calculation, remained potent enough to bias mismatch trial decision making towards 'match'. Moreover, the reduction in PASS+ card critical condition errors was modest (7%), remaining much higher than one would expect from if this were only a face matching task, and much higher than retailers, licencing authorities and policing would be likely to find acceptable.

In Experiment 3, we assessed whether a group of cashiers who were currently working in retail environments which required proof of age checks, would perform any better than a control sample with no such experience. While we found that the cashier group performed better than the controls in the critical condition, making 21% fewer incorrect decisions than controls, critical condition error rates remained far from perfect (34%). While it is the case that several studies have shown that experience is not associated with enhanced identity verification performance (e.g. Burton et al, 1999; White et al, 2014), there are now a number of recent studies which, in line with our current findings, do show that experienced identity checkers can outperform naïve samples, but without showing a meaningful 'step-change' in error rates (i.e. they still remain far from perfect; Phillips, Yates, Hu et al, 2018; Towler, White, & Kemp, 2017; White, Hahn, Hill, & O'Toole, 2015; White, Dunn, Schmid, & Kemp, 2015b).

In Experiment 3, we also report an unexpected trend between critical condition errors and cashier experience, which suggested that greater experience led to *greater* critical condition errors. Further experimentation will be needed to establish whether this relationship is reliable, as the reported association is non-significant and statistically weak. However, if the effect turns out to be robust, it may be that, given the time pressure cashiers are under, they focus on getting the age calculation correct as quickly as they can, paying less attention to the matching decision than their more inexperienced counterparts. Our findings on checking strategy are consistent with this explanation and the likelihood an age confirmation bias drives our critical condition effects. Across the three experiments, 69% of participants reported checking the date of birth first (i.e. the age verification task first) and then the face pairs, for the driving licence ID cards. However, it was the opposite strategy, checking the face pairs first and then performing the age calculation, that resulted in significantly fewer errors in the critical condition. It is important to note that this face-first strategy, while still not eliminating errors, provided a much bigger improvement in performance (26%) than simply switching to PASS cards (7%). Checking strategy was reported by participants after completion of the task, however, it is not clear from the current dataset whether they used this strategy on all trials. Therefore, future work in which strategy is explicitly manipulated will be needed to resolve any causal effects. Should they exist, there are clear benefits for staff training in checking ID.

We also note that we did not introduce time and social pressure into our design: most cashiers report being under pressure and somewhat uncomfortable at having to ask a customer for ID. Nor did we assess the effects of poor quality lighting on matching performance (i.e. the type of environment an ID checker may be placed in at a bar or nightclub; see Mileva & Hancock, 2019). In addition, the proportion of critical condition trials (i.e. fraud attacks) in our study may be larger than many cashiers would be presented with in daily work. Papesh and Goldinger (2014) have shown that reducing the frequency of face mismatches can lead to greater error rates, and the present data could be underestimating the likelihood of a fraud attack being successful. Finally, we note that some of our GFMT-long form faces would have looked older than 18 even in conditions in which the date of birth displayed on the card labelled them as underage for the purposes of buying alcohol. We now aim to investigate these factors in a series of follow up studies which will include a new set of face pairs from individuals in the 16-25 age range only, which will allow us to test the strength of this age confirmation bias with a fully age appropriate face set.

As an additional means of reducing error in this context, previously established means of improving unfamiliar face matching performance such as adding multiple photos of the bearer to the ID card (White, Burton, Jenkins & Kemp, 2014) or selecting individuals on the basis of their aptitude with faces (Bobak, Hancock, & Bate, 2016; Davis, Lander, Evans, & Jansari, 2016; Robertson, Noyes, Dowsett, Jenkins & Burton, 2016) should be tested in a task which also requires age verification to discern whether they further reduce the age confirmation bias.

Conclusions

In this study we investigated the potential for interactions between age and identity verification processes. Across three experiments we show that the confirmation that a “customer’s” ID card has been issued to an individual aged 18 or above results in a bias towards accepting mismatching face pairs (i.e. two difference people) as a match. Experienced cashiers showed fewer errors in this task, but performance was still far from perfect, and a face-first checking strategy was associated with improved performance to a greater extent than pre-verified PASS cards. These results emphasise the importance of understanding the details of unfamiliar face matching across a wide range of daily tasks involving an ID-check.

References

Alenezi, H. M., Bindemann, M., Fysh, M. C., & Johnston, R. A. (2015). Face matching in a long task: enforced rest and desk-switching cannot maintain identification accuracy. *PeerJ*, *3*, e1184.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive psychology*, *30*(3), 221-256.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Beattie, L., Walsh, D., McLaren, J., Biello, S. M., & White, D. (2016). Perceptual impairment in face identification with poor sleep. *Royal Society open science, 3*(10), 160321.

Bindemann, M., Fysh, M., Cross, K., & Watts, R. (2016). Matching faces against the clock. *i-Perception, 7*(5), 2041669516672219.

Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception, 40*(5), 625–627.

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

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLOS ONE*, *11*(2). e0148148.

Bobak, A. K., Hancock, P. J., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, *30(*1), 81-91.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42*(1), 286–291.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*(3), 243-248.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45. doi: 10.20982/tqmp.01.1.p042

Davis, J. P., Lander, K., Evans, R., & Jansari, A. (2016). Investigating predictors of superior face recognition ability in police super-recognisers. *Applied Cognitive Psychology, 30*(6), 827-840.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505.

Feng, X., & Burton, A. M. (2019). Identity Documents Bias Face Matching. *Perception, 48(12),* 1163-1174.

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British Journal of Psychology*, *109*(2), 219-231.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*(5), 577–596.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 11*(3), 211-222.

Klayman, J. (1995). Varieties of confirmation bias. *The Psychology of Learning and Motivation, 32*, 385–418.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339.

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own-and other-race faces. *Visual Cognition*, *21*(9-10), 1287–1305.

Mileva, V. R., & Hancock, P. J. (2019). A grey area: how does image hue affect unfamiliar face matching?. *Cognitive research: principles and implications, 4*(1), 27

McCaffery, J. M., & Burton, A. M. (2016). Passport checks: interactions between matching faces and biographical details. *Applied Cognitive Psychology, 30*(6), 925-933.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive research: principles and implications, 3*(1), 21. https://doi.org/10.1186/s41235-018-0112-9.

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own-and other-race faces. *Visual Cognition*, *21*(9-10), 1287-1305.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34*(4), 865–76.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied, 14*(4), 364.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29*(1), 85-95.

National Proof of Age Standards Scheme (PASS). (2018). Retrieved from http://www.pass-scheme.org.uk/ on 4th March 2019.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175-220.

Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive research: principles and implications*, *3*(1), 19.

Papesh, M. H., & Goldinger, S. D. (2014). Infrequent identity mismatches are frequently undetected. *Attention, Perception, & Psychophysics, 76*(5), 1335-1349.

Robertson, D. J., Black, J., Chamberlain, B., Megreya, A. M., & Davis, J. P. (2020). Super-recognisers show an advantage for other race face identification. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3608.

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by metropolitan police super-recognisers. *PloS one, 11*(2), e0150036.

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive Research: Principles and Implications*, *5*(1), 1-17.

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47.

UK Citizencard Scheme. (2019). Retrieved from https://www.citizencard.com/requirements-for-a-first-uk-id-card on 4th May 2019.

UK Home Office Report (2010). *New conditions for licensed premises in England and Wales: age verification and smaller measures. Retrieved from https://www.gov.uk/government/publications/new-conditions-for-licensed-premises-in-england-and-wales-age-verification-and-smaller-measures* on 4th May 2019.

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision research, 141*, 217-227.

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied, 20*(2), 166.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS one, 9*(8), e103510.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PLOS ONE, 10*(10). e0139827.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151292.

Wirth, B. E., & Carbon, C. C. (2017). An easy game for frauds? Effects of professional experience and time pressure on passport-matching performance. *Journal of Experimental Psychology: Applied, 23*(2), 138–157.

Figure Legends

Figure 1 An example of a 'critical condition' trial from Experiment 1 in which the 'customer' has presented an age appropriate driving licence, but in which there is mismatch between the face of the bearer and the face photo on the ID card (i.e. a fraud attack).

Figure 2 Graph showing the mean percentage response error rates found in Experiment 1, as a function of checking condition. Error bars show within-subjects standard error (Cousineau, 2005).

Figure 3 An example of a PASS card 'critical condition' trial from Experiment 2, no age calculation is required, and the faces mismatch. The 'Face Check Only' message appeared on all trials in this condition to re-inforce the pre-task instruction that only a face identity check was required for PASS identity cards.

Figure 4 Graph showing the mean percentage response error rates found in Experiment 2, as a function of checking condition. Error bars show within-subjects standard error (Cousineau, 2005).

Figure 5 Graph showing the mean percentage response error rates found in Experiment 3. Error bars show within-subjects standard error (Cousineau, 2005).

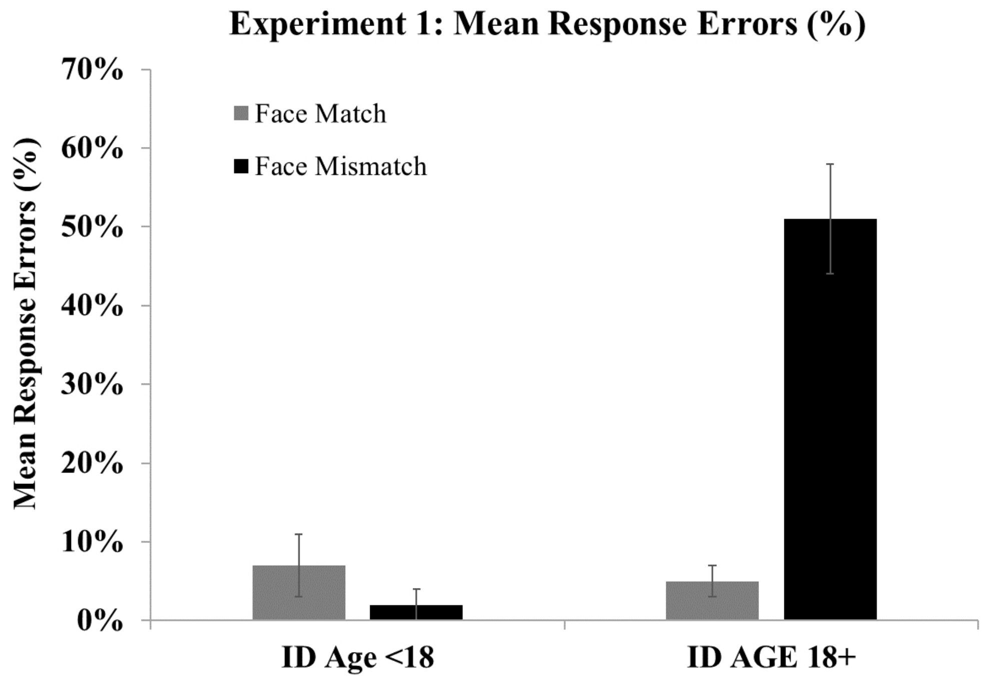
Supplementary Materials

Additional statistics for Experiment 3

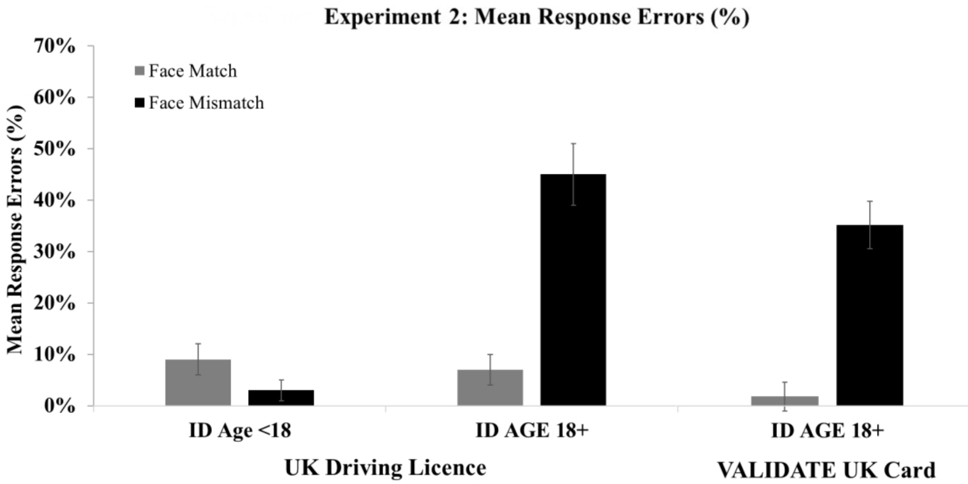
Main effect of age on card, $F(1, 34) = 70.02$, $p < .001$, $\eta_p^2 = .67$, main effect of face matching, $F(1, 34) = 63.97$, $p < .001$, $\eta_p^2 = .65$, main effect of group, $F(1, 34) = 6.01$, $p = .020$, $\eta_p^2 = .15$. Two-way interaction between face matching and group, $F(1, 34) = 6.05$, $p = .019$, $\eta_p^2 = .15$, two-way interaction between age on card and group, $F(1, 34) = 4.57$, $p =$

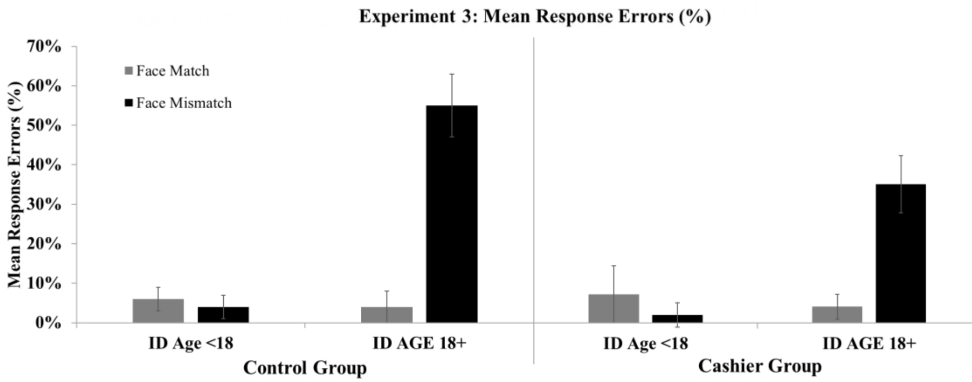
.040, η_p^2 = .12, two-way interaction between face matching and age on card, $F(1, 34)$ = 129.29, $p < .001$, η_p^2 = .79.


UK
PROVISIONAL DRIVING LICENCE
1. SNAPE
2. EMILY
3. 01.01.1992 UNITED KINGDOM
4a. 05.05.2014 4c. DVLA
4b. 04.05.2024
5. BYHKB65161HY5BK 56
7. E. SNAPE
8. 1 HIGH STREET
BRISTOL
HY5 6KB
9. HY/H/K5/K/B/K/Y/H/B/K
04.05.2024
DVLA LICENCE WSE
UOY PHOTOCARD
It is illegal to
sell alcoholic products
to anybody under
the age of 18


Experiment 1: Mean Response Errors (%)
Mean Response Errors (%)
70%
60%
50%
40%
30%
20%
10%
0%
Face Match
Face Mismatch
ID Age <18
ID AGE 18+


Face Check
Only
VALIDATE UK
Proof of Age Card
Name
DANIEL BENNETT
DoB
23 Sep 1992
2177 7127 7712 1771
PASS
18+
It is illegal to
sell alcoholic products
to anybody under
the age of 18


Experiment 2: Mean Response Errors (%)
Face Match
Face Mismatch
Mean Response Errors (%)
70%
60%
50%
40%
30%
20%
10%
0%
ID Age <18
ID AGE 18+
ID AGE 18+
UK Driving Licence
VALIDATE UK Card


Experiment 3: Mean Response Errors (%)
Mean Response Errors (%)
70%
60%
50%
40%
30%
20%
10%
0%
Face Match
Face Mismatch
ID Age <18
ID AGE 18+
Control Group
ID Age <18
ID AGE 18+
Cashier Group