Decision-Making Framing in Facial Image Comparison

Kristen A. Baker[1]*, Markus Bindemann. [1]

[1]University of Kent

* Corresponding author information: Kristen A. Baker, University of Kent, CT2 7NZ, Canterbury, Kent, United Kingdom. (email: k.baker-737@kent.ac.uk)

This paper is published in *Journal of Applied Research in Memory and Cognition*

Abstract

Unfamiliar face identification is a challenging task and is often considered to be perceptual in nature. This ignores the contribution of decision making in face identification. Across three experiments, we examined whether face-matching decisions would change depending on the framing of response options. In Experiment 1, participants completed an unfamiliar face matching task with the Same/Different, Same/Not-Same or Same/Proceed-to-Next-Trial response options. In Experiment 2, participants completed the task twice, once with the Same/Different response options and once with the Same/Different, Same/Not-Same or Same/Proceed-to-Next-Trial response options. In Experiment 3, participants completed the task when 50% or 80% of trials were matches and used the Same/Different or Same/Proceed-to-Next-Trial response options. Whereas we found no evidence that face identification decisions were influenced by the framing of response options, we found evidence of a robust base rate effect. Together these findings suggest that decision making in unfamiliar face identification is a complex and selective process.

Keywords:

unfamiliar face identification, face matching, face recognition, decision making, framing

General audience summary

Photographic identity documents are used in a many applied settings, such as when crossing international borders or purchasing age-restricted goods. However, facial image comparison is difficult for lay persons and professionals; this is because images of the same person can look very different and different people can look very similar. This premise has resulted in a focus on the perceptual difficulties in unfamiliar face identification; we used a novel approach and instead focused on decision-making aspect of this task. There is an abundance of evidence to suggest that the way in which questions, or the corresponding response options, are posed influences decision making. Here, we examine whether the way in which the response options are posed influenced decision making in unfamiliar face identification. Just as in these applied settings, participants determined whether pairs of faces were of the same person. Critically, they made these decisions with *neutral response options* (i.e., Same/Different), which should not bias participants towards making either response, or *biased response options* (e.g., Same/Not-Same; Same/Proceed-to-Next-Trial), which should bias responses towards making more “same” responses. Because in most applied settings it is much more likely that a person is presenting a valid photographic identity document, we also manipulated how often face pairs depicted images of the same person. Although participants’ decisions were not influenced by the framing of response options, they responded “same” more often when more of the trials depicted two images of the same identity. These results highlight the importance of understanding factors that influence face-matching decisions. While the framing of response options cannot be used to influence performance in applied settings, informing individuals of the frequency of matched versus mismatched documents might reduce the most egregious errors.

Decision-making framing in facial image comparison

Societies depend on person identification from photo-identity documents, a task known as facial image comparison in police and border security settings. Despite its wide-spread use, facial comparison is challenging as two photos of the same person can look very different and images of two people can look remarkably similar. In Psychology, this process is examined using *unfamiliar face matching* tasks, wherein participants make identity judgements for pairs of face images that depict the same person (matches) or different people (mismatches). This research shows that face matching is error prone for lay participants (Baker et al., 2023; Burton et al., 2010; Fysh & Bindemann, 2018), security professionals (Towler et al., 2019; White et al., 2014; Wirth & Carbon, 2017) and photographers, who understand variation in facial appearance (Baker & Mondloch, 2022).

As this challenge relies on similarity judgments of faces (Baker & Mondloch, 2019; Burton, 2013; Jenkins et al., 2011), it is often approached as a perceptual problem (Baker et al., 2023; Bindemann & Burton, 2021). However, there is evidence that it is also a decision-making problem, as the perceptual information in facial images can be dissociated from responses. Different observers often make distinct decisions to the same faces (Bindemann et al., 2012a; Fysh & Bindemann, 2018; Stantic et al., 2022) and also vary their identification decisions to the same faces over time (Alenezi & Bindemann, 2013; Alenezi et al., 2015; Bindemann et al., 2012b). This should not arise if these errors simply reflect the available *visual* information in faces and points to inconsistencies in how this information is used to make identification *decisions*. Contextual manipulations, such as the provision of example stimuli, also influence identification decisions even though visual information is held constant (Gentry & Bindemann, 2019; Papesh et al., 2018). Consequently, there have been calls to integrate decision making into models of face matching (Bindemann & Burton, 2021).

Decision-making models postulate that evidence is compared to a decision threshold (Ratcliff et al., 2016; Summerfield & Egner, 2014). For example, in experiments utilising *signal detection theory*, binary decisions depend on the available evidence and position of one's decision boundary or threshold (i.e., *criterion*; see Stanislaw & Todorov, 1999). A decision is reached when there is sufficient evidence relative to one's criterion, indicating that that the evidence was likely obtained from one distribution (an identity match) over another (a mismatch). Investigating criterion can provide a means of understanding decision making in face identification.

In face matching, individual differences in criterion placement are stable across time and tasks, indicating that observers engage consistent decision-making processes (Baker et al., 2023)—a finding that is robust across stimulus categories and measures (Aminoff et al., 2012; Baldassari et al., 2019). However, individuals shift their criterion in situations that are characterized by uncertainty (Lynn & Barrett, 2014; Lynn et al., 2015; Summerfield & Tsetsos, 2012), such as when identity matches and mismatches occur with different frequencies (unequal base rates) and different costs are associated with these responses (Stabile et al., 2024). Thus, criterion for these identification decisions can be influenced independently of the perceptual information conveyed by the faces.

Little is known about manipulations that influence face-matching decisions, such as whether the demands of this task must be changed directly or whether it is sufficient to change observers' task *concept*. There is evidence that changes to the task concept can influence face-matching decisions. In comparison to laboratory face-matching tasks, participants' decision-making boundaries (criterion) shift towards match responses in airport settings (Bindemann et al., 2022; Tummon et al., 2019) and when faces are embedded in photographic identity documents (Feng & Burton, 2019, 2021; McCaffery & Burton, 2016). It has been proposed that these effects might reflect participants' implicit expectations that most document bearers and their photographic identity documents are identity matches, even if they look quite different (Bindemann et al., 2022; Feng & Burton, 2019, 2021).

Given that expectations have been shown to influence perception in other domains (Bang & Rahnev, 2017; Haarsma et al., 2025; Stein & Peelen, 2015), it is possible that they might also affect face matching. For example, implicit expectation effects have been induced by changing the framing of response options in the perceptual discrimination of the orientation of Gabor patches (Summerfield & Koechlin, 2008). Observers classified two stimulus categories using A/B response options, which should not induce expectations that one category is more probable than the other, or A/Not-A response options, which could indicate that the ‘A’ category is more probable. Participants were faster to make ‘A’ decisions in the A/Not-A than the A/B condition (Summerfield & Koechlin, 2008). These effects also occur for decision criterion in recognition memory tasks when test questions are framed as ‘is the item old?’ compared to ‘is the item new?’ (Leynes, 2021; Mill & O’Connor, 2014).

Similar effects have been observed in the heuristics and decision-making literature. Individuals often respond differently to the same problem depending on how it is framed, despite evidence not differing between separate versions (Tversky & Kahneman, 1981; see Keren, 2010; Kühberger, 1998; Maule & Villejoubert, 2020). For example, framing a bag of chips as having ‘60% less salt’ compared to ‘40% salt’ can result in considerable differences in consumer behaviour. One explanation for such effects is that problems are represented in terms of gains and losses relative to a reference point, which can vary because of observers’ expectations (Tversky & Kahneman, 1981). Indeed, framing effects have even been adapted for use in other fields such as politics, environmental issues and health (Gong et al., 2013; Homar & Cvelbar, 2021; Lechler & de Vreese, 2011).

Considering the prevalence of framing effects, we examined whether this influences decision making in unfamiliar face matching, too. Across three experiments, participants were randomly assigned to a biased response option condition (Same/Not-Same; Same/Proceed-to-Next-Trial) or a neutral response option condition (Same/Different), following a similar format utilised in the perceptual decision-making literature (Summerfield & Koechlin, 2008). We examined how the framing of response options affected observers' sensitivity (d') to facial information, their decision-making boundaries (*criterion*), and the difference in response times for match and mismatch trials. Considering framing and expectation effects (Bang & Rahnev, 2017; Summerfield & Koechlin, 2008; Tversky & Kahneman, 1981), we hypothesised that d' would be unaffected by these response options but *criterion* would differ. Essentially, more face pairings would be classified as matches (liberal criterion) in the Same/Not-Same and Same/Proceed-to-Next-Trial conditions compared to the Same/Different condition. Based upon the findings of Summerfield and Koechlin (2008), we expected participants would respond faster on match than mismatch trials in the Same/Not-Same and Same/Proceed-to-Next-Trial conditions compared to the Same/Different condition. This is because an increased expectation of match trials might lead to an increased readiness or reduced evidence required for faster executions of 'same' responses.

Experiment 1

We investigated whether the framing of response options influences face-matching decisions. This will provide evidence as to whether these decisions are only influenced by external manipulations (Bobak et al., 2019; Feng & Burton, 2019, 2021) or also by observers' internal conceptualisation of this task. Participants completed response options conditions wherein they classified face pairs as Same/Different, Same/Not-Same or Same/Proceed-to-Next-Trial. As with previous research with simple visual stimuli such as Gabor patches (Bang & Rahnev, 2017; Summerfield & Koechlin, 2008), we hypothesised that more pairings would be classified as matches (liberal criterion), and that matches would be classified faster than mismatches in the Same/Not-Same and Same/Proceed-to-Next-Trial conditions compared to the Same/Different condition.

Method

Participants

Ninety white young adults (N_{female} = 55, M_{age} = 28.27 years, SD_{age} = 5.18) participated in this experiment. These participants were recruited via prolific.com (*n* = 82) for a small fee (£4.50) or from the School of Psychology at the University of Kent (*n* = 8) for course credit. This sample size was determined using a priori analyses in G*Power (Faul et al., 2007; Faul et al., 2009) for a one-way ANOVA with three levels (Same/Different, Same/Not-Same, Same/Proceed-to-Next-Trial), with power of 0.80, an *alpha* threshold of 0.05, and an effect size of 0.33 (a medium-large effect). Whereas Stabile et al. (2024) found a large effect in a within-subject design, our calculations are based on a medium-large effect, as the current experiment employed a between-subjects design. An additional three participants were tested, but excluded from the analyses because of failed attention checks (*n* = 3; see stimuli section) or failure to report the correct response options that they used during the task (*n* = 0; see Procedure section). All participants provided informed consent.

Stimuli

Participants completed the short version of the Kent Face Matching Test (KFMT), which comprises 20 match and 20 mismatch face pairings (Fysh & Bindemann, 2018). Each face pair consists of an image taken from student identity cards, at a size of 142 x 192 pixels with a screen resolution of 72 ppi, and a passport-style face portrait at a size of 283 x 332 pixels. A screen calibration (see Procedure) maintained the aspect ratio and sizes of all images across participants. In identity matches, two photographs of the same person were paired. In mismatches, photographs of two different people of similar appearance were combined.

In addition to these 40 face pairings, four further stimulus pairs were created to serve as attention checks. The aim of these attention checks was to create a set of face pairings that could be resolved easily. In two of these displays, identical face photographs were paired to serve as image matches. In the other two displays, a male and a female face were combined as sex mismatches. These pairs were created from additional trials of the KFMT-long form (Fysh & Bindemann, 2018). These trials were randomly intermixed with the test stimuli and presented at the same size.

Procedure

All participants were tested using an online platform (www.pavlovia.org). Prior to the face-matching task, a screen calibration was conducted to ensure that the stimuli were presented with a correct aspect ratio and size. Participants adjusted an image of an on-screen credit card to the dimensions of a physical credit card. Participants then completed 44 trials of the task (i.e., 40 KFMT trials and four attention-check trials). Participants were randomly assigned to one of three groups in which they were either asked to classify the face pairings as Same/Different, Same/Not-Same or Same/Proceed-to-Next-Trial; 30 participants were assigned to each condition. In the Same/Different condition, participants were informed that their 'task was to determine whether the pair of images were of the same person or of different people'. In the Same/Not-Same condition, participants were informed that their 'task was to determine whether the pair of images were of the same person or not of the same person'. In the Same/Proceed-to-Next-Trial condition, participants were informed that their 'task was to determine whether the pair of images were of the same person. If the trial did not contain two images of the same identity, [they were] to proceed to the next trial'.

Each trial comprised a pair of face images being presented in the centre of the screen. The face pair remained on screen until participants indicated their response. Participants were asked to press the ‘F’ key for ‘Same’ responses and the ‘J’ key for all other responses (i.e., Different, Not-Same, Proceed-to-Next-Trial). The response options remained on screen throughout to ensure that participants were always aware of the task at hand. Performance was not incentivized and feedback for the accuracy of responses was not provided. Trial order was randomised for each participant.

Immediately following completion of the task, participants were asked to report the response options that they used throughout the experiment. This enabled us to be more confident that any effects were the result of participants attending to the response options. This protocol was pre-registered on OSF (https://osf.io/fs5kd) and all corresponding data for each experiment are available online (https://osf.io/c6yqa/).

Results

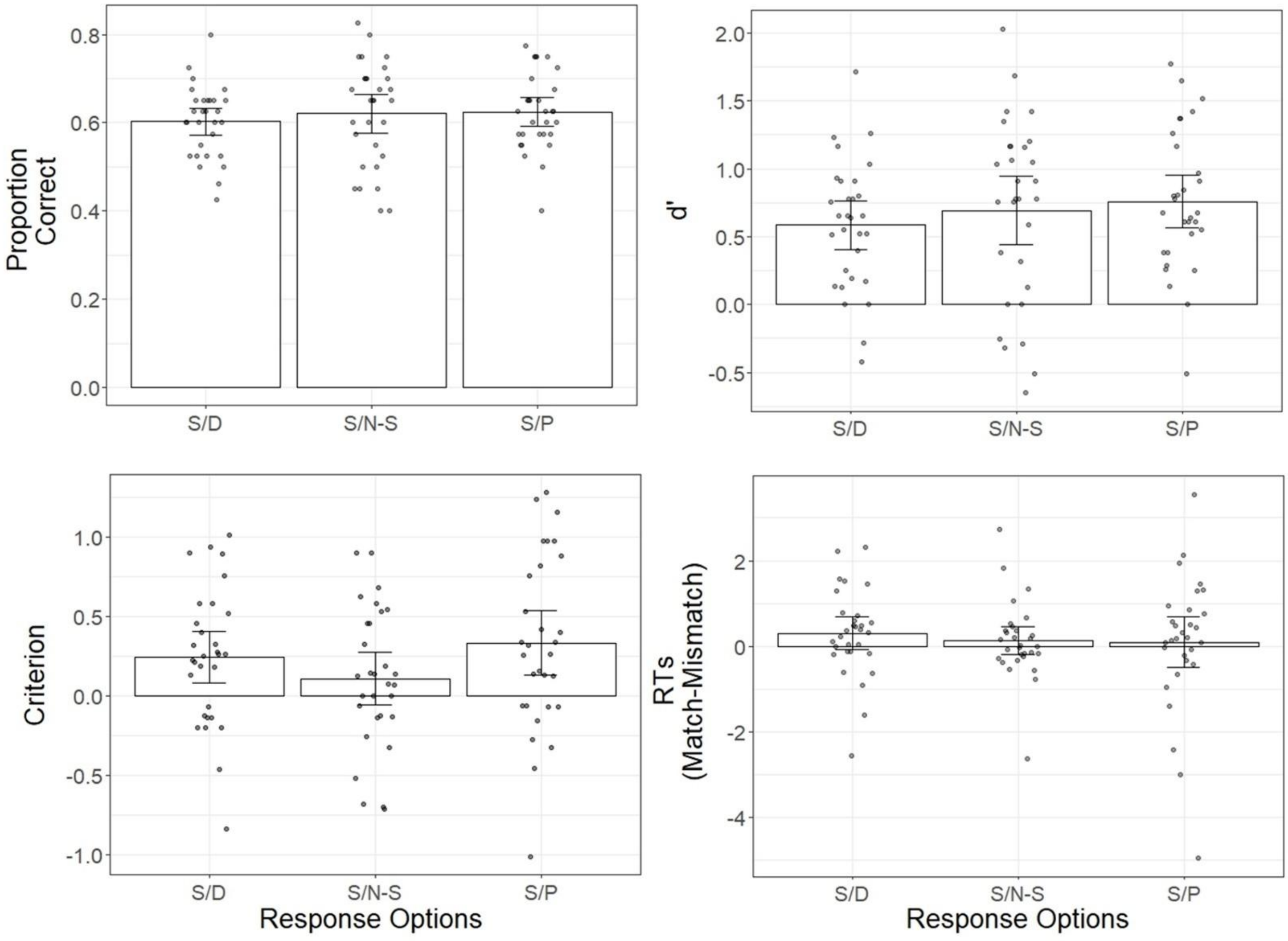
In a preliminary analysis, we analysed the proportion of correct responses for the three experimental conditions (Same/Different, Same/Not-Same, Same/Proceed-to-Next-Trial). This shows that just over 60% of face pairs were classified correctly and is comparable to the normative data for the short KFMT (66%; see Fysh & Bindemann, 2018). Here and throughout, all null effects were followed up with Bayesian analyses using JASP 0.18. We found no effect of response options on accuracy. These findings are reflected by a one-way ANOVA, $F(2, 87) = 0.42$, $p = 0.66$, $\eta^2 = 0.01$, $BF_{10} = 0.14$. All data are displayed in Figure 1 and summary statistics are also provided in the supplementary file S1. These data comprise all values, including separate means for match and mismatch accuracy and response times (RTs).

For our primary analyses, we converted these data into the signal detection measures of *d′* and *criterion* using hits (defined here as correct ‘same’ responses on match trials) and false alarms (defined here as incorrect ‘same’ responses on mismatch trials). In contrast to previous studies that have analysed match and mismatch accuracy independently (e.g., Bobak et al., 2023; Megreya & Bindemann, 2015; Ritchie & Burton, 2017), we adopted this approach specifically to investigate changes in decision-making *criterion* as a function of response-option framing. Participant responses on both match and mismatch trials are influenced by *criterion* and *d′*, and the signal detection approach makes it possible to separate these decision-making characteristics from sensitivity to facial information (Stanislaw & Todorov, 1999). For this purpose, hit and false alarm rates of 0 or 1 were corrected using $(n_{Signal\ or\ Noise\ trials} - 0.5)/n$ and $0.5/n_{Signal\ or\ Noise\ trials}$, respectively (Macmillan & Kaplan, 1985). Perfect performance would be reflected in $d' = 3.92$. A neutral criterion would be reflected as $c = 0$. Consistent with our hypothesis that the framing of response options should not affect sensitivity, we found that *d′* was comparable across all conditions, $F(2, 87) = 0.71$, $p = 0.49$, $\eta^2 = 0.02$, $BF_{10} = 0.17$. Contrary to our hypothesis that response options should affect *criterion*, we found that it was comparable across all conditions, $F(2, 87) = 1.70$, $p = 0.19$, $\eta^2 = 0.04$, $BF_{10} = 0.38$.

Finally, we analysed the RTs across conditions to investigate whether the framing of response options affected the speed with which face pairs were classified. To create this measure, we calculated each participant’s median RT (in seconds) using only the trials on which they were correct. This was done separately for match and mismatch trials. We then created a difference score by subtracting each participant’s mismatch RT from their match RT. Thus, positive values represent faster RTs for match trials, and negative values represent faster RTs for mismatch trials. This difference score was applied in all subsequent RT analyses. Contrary to our hypothesis that response options should affect the difference in speed with which participants responded to match and mismatch trials, the one-way ANOVA did not reveal an effect of response-option condition, $F(2, 87) = 0.25$, $p = 0.78$, $\eta^2 < 0.01$, $BF_{10} = 0.12$.

Figure 1

Effects of response options for all performance measures in Experiment 1.



Note. White, light grey and dark grey bars reflect participants' performance (proportion correct, *d', criterion* and RTs) when assigned to the Same/Different (S/D), Same/Not-Same (S/N-S) and Same/Proceed-to-Next-Trial(S/P) response-option conditions, respectively. Individual data points are displayed for each participant. Error bars represent 95% CIs.

Discussion

This experiment examined whether the framing of response options affects face-matching decisions. In contrast to previous studies with simple visual stimuli (Summerfield & Koechlin, 2008), the framing of response options did not affect face identification. This was evident across accuracy and sensitivity —and most importantly, RTs and decision-making criterion. This is in contrast with findings from face-matching studies in which other variables are manipulated (e.g., match versus mismatch base rates) and produce a reliable criterion shift (Stabile et al., 2024). In Experiment 2, we sought to confirm the absence of response framing effects in face matching by controlling for individual differences. It is also possible that Experiment 1 failed to detect a difference between response-option conditions as sample size estimates were based on medium-large effects. In Experiment 2, power was therefore increased in case response-option framing only produces small effects on face matching.

Experiment 2

In Experiment 1, response conditions were administered on a between-subjects basis. However, face matching is characterised by substantial individual differences in accuracy and criterion (Baker et al., 2023; Baker & Mondloch, 2023; Bobak et al., 2023; Fysh & Bindemann, 2018). The measures of dispersion for Experiment 1 indicate a broad range in individual differences. Thus, individual differences in performance might have masked the effect of response options. Experiment 2 examined this possibility by having participants complete the KFMT twice: Once in a control condition, wherein participants responded to stimuli using the Same/Different response options, and once in an experimental condition, wherein participants responded using one of the response conditions from Experiment 1. As within-subjects designs control for individual variability between conditions, this is an effective method for isolating response-option effects. We hypothesised that more face pairings would be classified as identity matches (liberal criterion), and that participants would classify matches faster than mismatches in the Same/Not-Same and Same/Proceed-to-Next-Trial conditions, compared to the Same/Different condition—but only during the experimental condition. These interactions would provide evidence that response options affect face-identification decisions. In addition, sample size was also increased substantially in case response options effects are smaller than anticipated.

Method

Participants

One hundred and eighty-nine white young adults (N_{Female} = 115, $N_{\text{Non-binary}}$ = 1, N_{Male} = 73; M_{age} = 25.67 years, SD_{age} = 3.01) participated in this experiment for a small fee (£4.50) and were recruited via www.prolific.com. This sample size was determined using a priori analyses in G*Power for a 2 (control condition, experimental condition) x 3 (Same/Different, Same/Not-Same, Same/Proceed-to-Next-Trial) mixed-factor ANOVA with power of 0.80, an *alpha* threshold of 0.05, and an effect size of 0.1 (a small effect). We opted to estimate the sample size based on a small effect to ensure the null finding from Experiment 1 does not reflect a Type 2 error. An additional nine participants were tested, but were excluded from the analyses because of failed attention checks (n = 3), or failure to report the correct response options that they used during the tasks (n = 6). All participants provided informed consent.

Stimuli and Procedure

The stimuli and procedure were the same as in Experiment 1, except for the following changes. In Experiment 2, all participants completed the KFMT-short and the attention check trials twice, once as a control version and once as an experimental version. In the control version, all participants responded using the Same/Different response options. In the experimental version, participants were randomly assigned to one of the three response options (Same/Different, Same/Not-Same or Same/Proceed-to-Next-Trial); 63 participants were assigned to each condition. In addition to randomising trial order, the order of the control and experimental versions was counterbalanced across participants. This protocol was pre-registered on OSF (https://osf.io/jf9rd).

Results

We analysed the data following the same protocols as in Experiment 1. All data are displayed in Figure 2 and summary statistics are also provided in the supplementary file S2. Consistent with Experiment 1, there was no effect of response options on accuracy. These findings are reflected in our preliminary analysis, wherein the 2 (control version, experimental version) x 3 (Same/Different, Same/Not-Same, Same/Proceed-to-Next-Trial) mixed-factor ANOVA showed that participants' accuracy was not influenced by task version, $F(1, 186) = 2.51$ $p = 0.11$, $\eta_p^2 = 0.01$, $BF_{10} = 0.37$; response options, $F(2, 186) = 0.10$, $p = 0.91$, $\eta_p^2 < 0.01$, $BF_{10} = 0.13$; or an interaction between these factors, $F(2, 186) = 0.32$, $p = 0.73$, $\eta_p^2 < 0.01$, $BF_{include} = 0.07$.

We then converted the data into the signal detection measures of *d′* and *criterion* for our primary analyses. Consistent with Experiment 1, and our hypothesis that response-option framing would not affect sensitivity, the 2 (version) x 3 (response option) mixed-factor ANOVA for *d′* revealed that there was not a main effect of task version, $F(1, 186) = 2.78$, $p = 0.10$, $\eta_p^2 = 0.02$, $BF_{10} = 0.43$; or of response options, $F(2, 186) = 0.12$, $p = 0.89$, $\eta_p^2 < 0.01$, $BF_{10} = 0.12$; and no interaction between these factors, $F(2, 186) = 0.54$, $p = 0.59$, $\eta_p^2 = 0.01$, $BF_{include} = 0.08$. Also consistent with Experiment 1, we found no support for our hypothesis that the framing of response options would affect *criterion*. The 2 (version) x 3 (response option) mixed-factor ANOVA for *criterion* revealed that there was not a main effect of task version, $F(1, 186) = 0.22$, $p = 0.64$, $\eta_p^2 < 0.01$, $BF_{10} = 0.12$; or of response options, $F(2, 186)$

= 0.95, p = 0.39, η_p^2 = 0.01, BF_{10} = 0.25; and no interaction between these factors, F(2, 186) = 0.04, p = 0.96, η_p^2 < 0.01, $BF_{include}$ = 0.05.

Finally, we analysed RTs to determine whether framing affected the difference in speed with which participants classified match and mismatch trials[1]. Consistent with Experiment 1, we found no support for our hypothesis that the framing of response options would affect the speed with which participants responded to match and mismatch trials. The 2 (version) x 3 (response option) mixed-factor ANOVA for the difference in RTs revealed that there was not a main effect of task version, F(1, 185) = 2.73, p = 0.10, η_p^2 = 0.01, BF_{10} = 0.38; or of response options, F(2, 185) = 0.36, p = 0.69, η_p^2 < 0.01, BF_{10} = 0.08; and no interaction between these factors, F(2, 185) = 0.38, p = 0.69, η_p^2 < 0.01, $BF_{include}$ = 0.07.

Exploratory Analyses

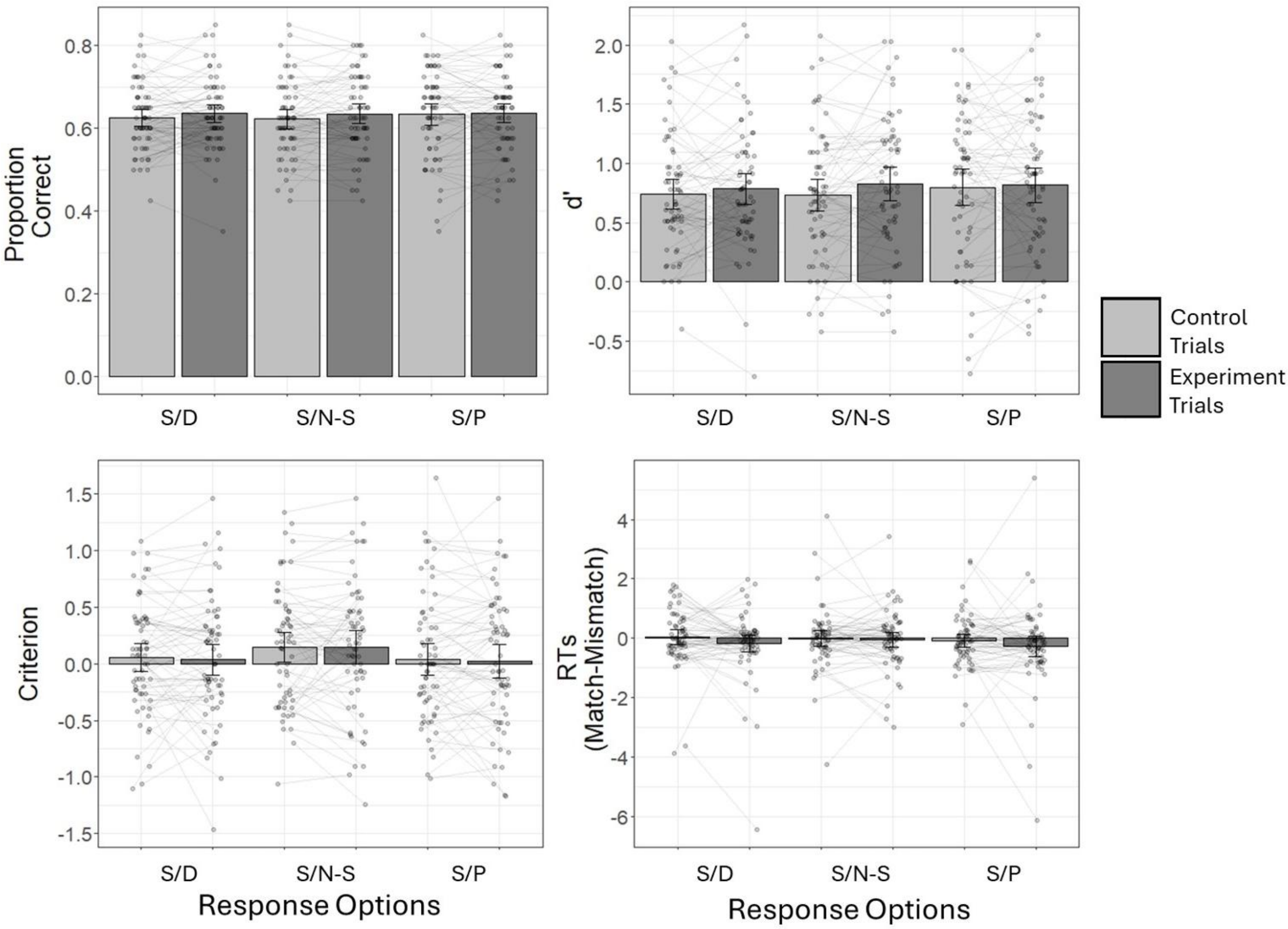
While the experimental and control conditions were counterbalanced to mitigate against practice effects, this might have attenuated the individual differences (for a brief discussion see Baker et al., 2023; Dale & Arnell, 2014). To confirm that this was not the reason for the null effect between response-option conditions, we conducted a series of exploratory Pearson correlations to investigate the relationship between experimental and control trials for each response-option group. For d', these correlations were significant in the Same/Different, r(63) = 0.68, p < 0.001; Same/Not-Same, r(63) = 0.72, p < 0.001; and Same/Proceed-to-Next-Trial conditions, r(63) = 0.70, p < 0.001. This finding holds true when this correlation was analysed across all conditions, r(189) = 0.70, p < 0.001. Similarly,

[1] One participant was excluded from all RT analyses as, despite passing attention checks, they were not correct on any match trials in the control condition. Thus, we could not calculate their score in the control task and opted to not use their RT data.

correlations between the experimental and control conditions were observed for *criterion* in the Same/Different, $r(63) = 0.67$, $p < 0.001$; Same/Not-Same, $r(63) = 0.75$, $p < 0.001$; and Same/Proceed-to-Next-Trial conditions, $r(63) = 0.72$, $p < 0.001$. This finding holds true when this correlation was analysed across all conditions, $r(189) = 0.72$ $p <0.001$. These results are consistent with previous reports of the stability of d' and criterion when face-matching performance is measured across different trials and testing sessions (Baker et al., 2023).

Figure 2

Effects of response options for control and experimental conditions in Experiment 2.



Note. Participants' performance in the Same/Different, Same/Not-Same and Same/Proceed-to-Next-Trial response-option conditions is reflected by the light grey (Control trials) and dark grey bars (Experimental trials), respectively. Individual data points are displayed for each participant, and connected for the control and experimental trials with a line. Error bars reflect 95% CIs.

Discussion

This experiment demonstrates that the framing of response options does not affect unfamiliar face matching even after accounting for individual differences in performance. This was observed across accuracy, sensitivity and, most importantly, decision-making criterion and RTs. As the data from Experiment 2 were collected under more tightly controlled conditions that required participants to complete the face-matching task twice, this strengthens the findings from Experiment 1. However, the question also arises whether this design induced repetition or practice effects that undermine response-option effects. We note that many face-matching studies demonstrate no improvement in face-matching accuracy with such repetition or practice (e.g., Alenezi & Bindemann, 2013; White et al., 2013; White et al., 2014; Towler et al., 2019) *and* we controlled for such effects by counterbalancing the response-option tasks. Moreover, to examine whether our design might have reduced variability in individual differences, we also examined the consistency of performance across these repetitions. This revealed strong correlations in performance between the control and experimental conditions, consistent with previous reports (e.g., Baker et al., 2023). This indicates that task repetition and counter-balancing did not limit our ability to statistically control for individual differences. Collectively, Experiments 1 and 2 therefore suggest that response-option framing does not affect unfamiliar face-matching decisions. This is striking as other variables show reliable criterion shifts in unfamiliar face matching, such as manipulations to match and mismatch base rates (Stabile et al., 2024). In Experiment 3, we sought to confirm the absence of response-option framing effects in unfamiliar face matching, by combining this with a base rate manipulation.

Experiment 3

Experiments 1 and 2 suggest that response-option framing does not influence decision-making criterion in face matching. This raises the possibility that criterion shifts might not be possible under these conditions. In contrast, base rate manipulations are effective at inducing criterion shifts in face matching (Stabile et al., 2024). To better understand the differences in base rates and response options, both factors were manipulated in Experiment 3. Participants were randomly assigned to a neutral base rate condition, in which 50% of trials were matches, or a biased base rate condition, in which 80% of trials were matches. Participants were also randomly assigned to use the Same/Different or Same/Proceed-to-Next-Trial response options. This allowed us to examine whether response framing and base rates exert separable effects on face-matching decisions. We hypothesized that more face pairs would be classified as identity matches (liberal criterion) when 80% compared to 50% of trials were matches. If base rate manipulations promote *criterion* differences, even if response options do not, then this strengthens the case that framing does not influence criterion in a comparable way.

Method

Participants

Two-hundred and eight white young adults (N_{Female} = 124, $N_{\text{Non-binary}}$ = 6, N_{Male} = 78; M_{Age} = 25.82, SD_{Age} = 3.15) participated in this experiment for a small fee (£1.20) and were recruited via www.prolific.com. This sample size was determined using a priori analyses in G*Power (Faul et al., 2007; Faul et al., 2009) for a 2 (base rate) x 2 (response options) ANOVA, with power of 0.80, an alpha threshold of 0.05, and an effect size of 0.195 (a small-medium effect). A small-medium effect allowed us to account for our use of a between-subjects design, when previous base rate effects were established in a within-subjects design by Stabile and colleagues (2024), potential unexpected effects from response options, and to ensure equal sample sizes across conditions. An additional 26 participants were tested, but excluded from the analyses because of failed attention checks (n = 2) or failure to report the correct response options and/or base rates upon task completion (n = 24). All participants provided informed consent.

Stimuli and Procedure

The stimuli were the same as in Experiment 1, except for the following changes. To create the base rate manipulation, the number of test trials was increased from 40 to 100. Thus, all participants completed the 40 trials from the KFMT short and an additional 60 trials. To add the additional 30 match trials for the equal base rate condition and 60 match trials for the unequal base rate condition, additional face pairs from the long version of the KFMT were added to the stimulus set. However, the long version of the KFMT only provides a limited number of mismatch trials. To create the additional mismatch trials that were required for the equal base rate condition, a further 30 mismatches were constructed from other unused match trials in the long version of the KFMT. For this purpose, one image from each of two unused match trials were combined . In this manner, 30 additional mismatches were created from 60 unused match trials. In a short pilot study, we ensured that the new mismatch trials did not differ in difficulty (M = 0.74; SD = 0.16) than those originally included in the KFMT short (M = 0.67, SD = 0.20), $t(14)$ = 1.59, p = 0.13. These pairs were chosen by one of the experimenters based on physical similarity of the models and maintained the same image appearance and size as in the original stimuli. None of the face identities were repeated across trials.

The procedure was the same as in Experiment 1, except for the following changes. In Experiment 3, participants were randomly assigned to a response-option condition (Same/Different, Same/Proceed-to-Next-Trial) and to a base rate condition, wherein participants completed the KFMT with either 50% match trials (equal match/mismatch base rates) or 80% match trials (unequal base rates). There were 52 participants assigned to each condition. Participants were informed about the base rate of the trials and the response options that they were to use. At the end of the task, to ensure that participants were cognisant of the response options and base rates that they were to be using, they were asked to report both. This protocol was pre-registered on OSF (https://osf.io/m9pdw).

Results

The data were analysed following the same protocols as in Experiments 1 and 2. For all analyses we only considered performance on the 40 KFMT short trials, as this allowed for comparison of performance on the same trials regardless of base rate condition. These data are displayed in Figure 3 and summary statistics are also provided in supplement S3. Consistent with Experiments 1-2, there was no effect of response options on accuracy and no effect of base rate on accuracy. These findings are reflected in our preliminary analysis, wherein the 2 (base rate) x 2 (response options) ANOVA showed that participants' accuracy was not influenced by response options, $F(1, 204) = 0.12$, $p = 0.73$, $\eta_p^2 < 0.01$, $BF_{10} = 0.32$; base rate, $F(1, 204) = 1.43$, $p = 0.23$, $\eta_p^2 < 0.01$, $BF_{10} = 0.16$; or an interaction between these factors, $F(1, 204) = 3.93$, $p = 0.07$, $\eta_p^2 = 0.02$, $BF_{include} = 0.11$.

We then converted these data into the signal detection measures of *d′* and *criterion* for our primary analyses. Consistent with Experiments 1-2 and our hypothesis that response-option framing and base rates would not affect *d′*, a 2 (base rate) x 2 (response options) ANOVA revealed that there was no main effect of response options, $F(1, 204) = 0.10$, $p = 0.76$, $\eta_p^2 < 0.01$, $BF_{10} = 0.16$; or of base rates, $F(1, 204) = 0.22$, $p = 0.64$, $\eta_p^2 < 0.01$, $BF_{10} = 0.17$; and no interaction between these factors, $F(1, 204) = 3.42$, $p = 0.07$, $\eta_p^2 = 0.02$, $BF_{include} = 0.07$). Also consistent with Experiments 1-2, the 2 (base rate) x 2 (response options)

ANOVA revealed no effect of response options on participants' *criterion*, $F(1, 204) = 0.95$, $p = 0.33$, $\eta_p^2 < 0.01$, $BF_{10} = 0.20$. In contrast to the findings for response-option framing, but consistent with our hypothesis that base rates would affect *criterion*, there was an effect of base rate, $F(1, 204) = 117.81$, $p < 0.001$, $\eta_p^2 = 0.37$, $BF_{10} = 7.31 \times 10^{+18}$. Participants were more liberal in their criterion, and therefore made more match responses, when 80% compared to 50% of trials were identity matches. No interaction of response options and base rate was found, $F(1, 204) = 0.10$, $p = 0.75$, $\eta_p^2 < 0.01$, $BF_{include} = 0.16$.

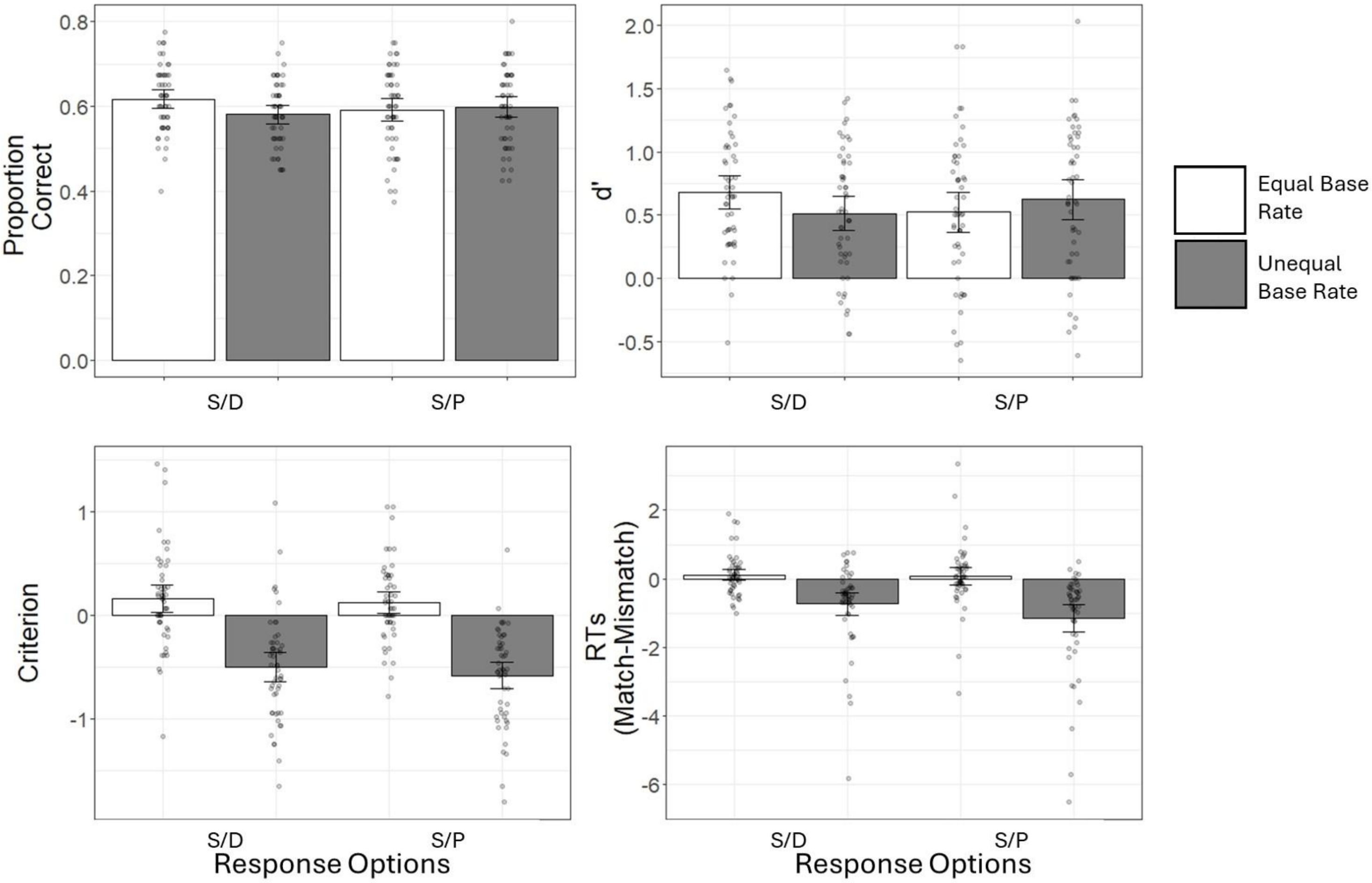
In addition to the pre-registered analysis, and to ensure that the absence of a response-options effect is not limited to the 40 trials that were consistently applied across conditions, we also analysed the full set of 100 trials of Experiment 3. The accuracy for the full set of trials was not influenced by response options, $F(1, 204) = 1.00$, $p = 0.32$, $\eta_p^2 < 0.01$, $BF_{10} = 0.21$; but was influenced by base rate, $F(1, 204) = 73.43$, $p < 0.001$, $\eta_p^2 = 0.26$, $BF_{10} = 1.49\text{x}10^{+12}$. The interaction between these factors failed to reach significance, $F(1, 204) = 3.75$, $p = 0.05$, $\eta_p^2 = 0.02$, $BF_{Include} = 0.86$. For *d′,* neither the effect of response options, $F(1, 204) = 0.21$, $p = 0.64$, $\eta_p^2 = 0.001$, $BF_{10} = 0.17$; base rate, $F(1, 204) = 0.60$, $p = 0.44$, $\eta p2 = 0.003$, $BF_{10} = 0.20$; or the interaction, $F(1, 204) = 2.07$, $p = 0.15$, $\eta p2 = 0.01$, $BF_{Include} = 0.05$; were significant. For *criterion*, while response options had no influence, $F(1, 204) = 2.22$, $p = 0.14$, $\eta p2 = 0.01$, $BF_{10} = 0.28$; base rates did, $F(1, 204) = 140.65$, $p < 0.001$, $\eta p2 = 0.41$, $BF_{10} = 5.68 \times 10^{+21}$. However, the interaction between these factors was not significant, $F(1, 204) = 0.01$, $p = 0.91$, $\eta p2 < 0.01$, $BF_{Include} = 0.25$.

Although the RT analysis for Experiment 3 was not included in our pre-registration, for completeness this is also presented here[2]. Consistent with Experiments 1-2, the 2 (base rate) x 2 (response options) ANOVA revealed no effect of response options on the difference in RTs between match and mismatch trials, $F(1, 202) = 2.26$, $p = 0.13$, $\eta_p^2 = 0.01$, $BF_{10} = 0.33$. In contrast to the findings for response-option framing, but consistent with our findings for *criterion*, there was an effect of base rate, $F(1, 202) = 46.56$, $p < 0.001$, $\eta_p^2 = 0.19$, $BF_{10} = 5.79 \times 10^{+7}$. Participants showed faster RTs for match compared to mismatch trials when 80% of trials were identity matches. No interaction of response options and base rate was found, $F(1, 202) = 1.53$, $p < 0.001$, $\eta_p^2 = 0.008$, $BF_{10} = 0.47$.

[2] Two participants, despite passing attention checks, were inaccurate on all mismatch or match trials and were therefore excluded from all RT analyses.

Figure 3

Effects of response options and base rates in Experiment 3.



Note. Participants' performance in the Same/Different, and Same/Proceed-to-Next-Trial response-option conditions is reflected by the white (Equal Base Rate) and dark grey bars (Unequal Base Rate), respectively. Individual data points are displayed for each participant. Error bars reflect 95% CIs.

Exploratory Analyses

To provide further evidence that there is no effect of response-option framing on accuracy, *d′* and *criterion*, we collapsed across the data in Experiments 1-3 and analysed the Same/Different and Same/Proceed-to-Next-Trial conditions. To ensure the data were comparable across studies, we only analysed participants from the equal base rate condition from Experiment 3, and only included the data from the first task participants completed in Experiment 2. This yielded a total sample of 290 participants. The cross-subject means of these data are shown in supplementary file S4. We then conducted t-tests for accuracy, *d′* and *criterion*. As in Experiments 1-3, there was no effect of response options on accuracy, $t(288) = 0.44$, $p = 0.66$, $d = 0.05$, $BF_{10} = 0.14$; on *d′*, $t\ (288) = 0.12$, $p = 0.91$, $d = 0.01$, $BF_{10} = 0.13$; or *criterion*, $t\ (288) = 0.05$, $p = 0.96$, $d < 0.01$, $BF_{10} = 0.13$.

Discussion

Experiment 3 demonstrates a clear effect of base rate. Relative to the equal base rate condition, participants in the unequal base rate condition made more ‘same’ than ‘different’ responses, reflecting a liberal decision-making criterion. This occurred without the provision of feedback or incentives—extending the previous literature (Stabile et al., 2024). In contrast, the framing of response options did not influence face-matching decisions across accuracy, sensitivity and–most importantly, criterion. As is evident in a range of studies, 40 trials are generally sufficient for establishing criterion effects in face matching (e.g., Stabile et al., 2024; Feng & Burton, 2019, 2021; McCaffery & Burton, 2016) and also revealed large base rate effects here. These results are corroborated by RTs, which were influenced by base rates, but not response options. Two exploratory analyses, across the full set of 100 trials in Experiment 3 and by combining data from 290 participants across all three experiments, provided further evidence that there was no effect of response options when power is increased substantially. Taken together, this provides strong and consistent evidence that face-matching decisions are malleable in response to base rates, but not the framing of response options.

General Discussion

Although unfamiliar face matching is often approached as a perceptual task, there is evidence that it is also a decision-making problem (Alenezi & Bindemann, 2013; Bindemann et al., 2022; Feng & Burton, 2019, 2021; Stabile et al., 2024). We therefore examined whether response-option framing influences face-identification decisions. Across observers' accuracy, sensitivity, RT, and decision-making criterion, we found no evidence that response-option framing influences face-identification decisions. This pattern persisted after controlling for individual differences in performance and when power was increased by combining data from 290 participants across experiments.

The absence of a response-option effect is surprising considering the well-established nature of framing effects (Tversky & Kahneman, 1981; see Keren, 2011; Levin et al., 1998; Maule & Villejouber, 2007), and evidence that question and response-option framing alter performance (Lynes, 2021; Mill & O'Connor, 2014; Summerfield & Koechlin, 2008). Importantly, the absence of a response-option effect is also striking given the large effect of base rate in Experiment 3—a finding that, even though incentives and feedback were not used, is consistent with previous experiments (Kantner et al., 2015; Lynn & Barrett, 2014; Stabile et al., 2024). Therefore, despite providing conceptually similar information, such as the predominance of match trials, these findings suggest that response options and base rates exert distinct effects on unfamiliar face-matching decisions.

Framing effects are thought to be the result of a combination of how the question is *created* and how the question is *interpreted* (Keren, 2011; Levin et al., 1998; Maule &

Villejouber, 2007). Framing effects are reduced when individuals are required to deliberate more or to justify their responses (Levin et al., 1998; Maule & Villejouber, 2007). Similarly, the framing of response options on perceptual decisions of lower-level stimuli induces expectations about a stimulus' probability (Summerfield & Koechlin, 2008). We report evidence against this expectation induction in unfamiliar face-matching decisions[3]. This is likely driven by inherent expectations of the proportion of match and mismatch trials, which are likely anticipated to occur with comparable frequency. This suggests that expectation effects in unfamiliar face matching, such as those in airport settings (Bindemann et al., 2022; Tummon et al., 2019) or when stimuli are imbedded in photographic identity documents (Feng & Burton, 2019, 2021; McCaffery & Burton, 2016), can override these baseline expectations.

Unlike the framing of response options, base rate information provides unambiguous information that leaves little room for interpretation. Indeed, individuals typically shift their criterion more optimally for unequal base rates than costs/payoffs (Bohil & Maddox, 2001; Kantner et al., 2015; Maddox & Bohil, 2005; Stabile et al., 2024). This occurs because individuals would rather risk making a more costly error than be inaccurate (Bohil & Maddox, 2001; Kantner et al., 2015; Maddox & Bohil, 2005), and shifting criterion increases accuracy when base rates are unequal, but does not aid in accuracy when payoffs/costs are unequal. Evidence that individuals would rather rely on uninformative perceptual evidence than risk being incorrect because of relying on a decision rule is shown in recognition

[3] If we compare *criterion* in the Same/Different and Same/Proceed-to-Next-Trial conditions when the 50% base rate is explicit in Experiment 3 versus not stated in Experiment 1, it is clear that individuals did not interpret information from the response options as the probability of a match trials as shown by the null interaction ($p = 0.37$).

memory experiments (Miller & Kantner, 2020). For example, participants were asked to identify previously studied suspicious individuals, and were informed of the *vital* importance of either not missing targets, or not mistakenly identifying non-targets (decision rule; Kantner et al., 2015). Even participants who did not complete the study phase did not rely on the decision rule, and did not adopt the optimal criterion. This might, in combination with participants' expectations, explain why the framing of response options did not affect unfamiliar face-matching decisions. Relying on the framing of response options would have required participants to wilfully sacrifice accuracy.

Together, these findings indicate that decision making in face identification is a complex and selective process. This complexity is evident in that individuals wilfully use perceptual evidence, such as the appearance of eyebrows and eyes (Abudarham & Yovel, 2016; Abudarham et al., 2019; Tyler et al., 2022), and non-face information, such as base rates (Stabile et al., 2024), to inform their decisions. Other studies suggest that directing attention to facial marks/moles improves performance on face-matching tasks (Fysh & Bindemann, 2022; Towler et al., 2017, 2021). We showed that the evidence used to make these decisions extends beyond perceptual facial information and is more complicated than previously shown. This process is highly selective, as not all external information (e.g., response options, base rates) was weighted equally during face-identification decisions. This might reflect that individuals are both accumulating evidence during their face-matching decisions, and weighting the strength of this information based on its evidentiary value. This possibility is consistent with decision-making models, as evidence weighting is an aspect of the evidence-accumulation process (Lindow & Betsch, 2018; Ratcliff et al., 2016; Summerfield & Egner, 2014). Indeed, much like there are age-related changes in face identification (Baker et al., 2017; Laurence & Mondloch, 2016; Matthews et al., 2022), there are also age-related changes in the weighting of information during evidence accumulation

(Lindow & Betsch, 2018)—a similarity that might explain some of the variance in unfamiliar face identification.

The studies reported here have several theoretical implications. Face identification is often only considered as if it were a perceptual problem. Most theories and studies fall in line with this assumption (Bruce & Young, 1986; Valentine, 1991; O'Toole et al., 2018). However, there is a growing body of literature suggesting a role for decision making in face identification (Alenezi & Bindemann, 2013; Baker et al., 2023; Stabile et al., 2024). Our findings add to this literature by showing the complexity and selectivity of face-identification decisions. We suggest that current conceptualisations of face identification need to be updated to account for decision making, including its complexity, selectivity and the use of non-face information.

Acknowledgements: Funding from the Royal Society's Newton International Fellowship Scheme (# 221754) supported this research. We thank participants for their time in completing these experiments.

References

Abudarham, N., & Yovel, G. (2016). Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision*, *16*(3), 40-40. https://doi.org/10.1167/16.3.40

Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, *182*, 73-83. https://doi.org/10.1016/j.cognition.2018.09.002

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735-753. https://doi.org/10.1002/acp.2968

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. https://doi.org/10.7717/peerj.1184

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., Grafton, S. T.,& Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition*, *40*(7), 1016–1030. https://doi.org/10.3758/s13421-012-0204-6

Baker, K. A., & Bindemann, M. (2024). Decision Making in Face Identification. Retrieved from https://osf.io/c6yqa/.

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30. https://doi.org/10.1016/j.cognition.2016.12.012

Baker, K. A., & Mondloch, C. J. (2019). Two sides of face learning: Improving between-identity discrimination while tolerating more within-person variability in appearance. *Perception*, *48*(11), 1124-1145. https://doi.org/10.1177/0301006619867862

Baker, K. A., & Mondloch, C. J. (2022). Picture this: Photographers no better than controls for recognizing unfamiliar faces. *Perception, 51*(8), 591-595. https://doi.org/10.1177/03010066221098727

Baker, K. A., & Mondloch, C. J. (2023). Unfamiliar face matching ability predicts the slope of face learning. *Scientific Reports*, *13*(1), 5248. https://doi.org/10.1038/s41598-023-32244-w

Baker, K. A., Stabile, V. J., & Mondloch, C. J. (2023). Stable individual differences in unfamiliar face identification: Evidence from simultaneous and sequential matching tasks. *Cognition*, *232*, 105333. https://doi.org/10.1016/j.cognition.2022.105333

Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). The importance of decision bias for predicting eyewitness lineup choices: toward a lineup skills test. *Cognitive Research: Principles and Implications*, *4*, 1-13. https://doi.org/10.1186/s41235-018-0150-3

Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific reports*, *7*(1), 17072. https://doi.org/10.1038/s41598-017-16885-2

Bindemann, M., Avetisyan, M., & Rakow, T. (2012a). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied, 18*(3), 277. https://doi.org/10.1037/a0029635

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012b). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition, 1*(2), 96-103. https://doi.org/10.1016/j.jarmac.2012.02.001

Bindemann, M., & Burton, A. M. (2021). Steps towards a cognitive theory of unfamiliar face matching. *Forensic face matching: Research and practice*, 38-61.

Bindemann, M., Fysh, M. C., Trifonova, I. V., Allen, J., McCall, C., & Burton, A. M. (2022). Face identification in the laboratory and in virtual worlds. *Journal of Applied Research in Memory and Cognition, 11*(1), 120. https://psycnet.apa.org/doi/10.1037/mac0000050

Bobak, A. K., Jones, A. L., Hilker, Z., Mestry, N., Bate, S., & Hancock, P. J. (2023). Data-driven studies in face identity processing rely on the quality of the tests and data sets. *Cortex, 166*, 348-364. https://doi.org/10.1016/j.cortex.2023.05.018

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). A grey area: how does image hue affect unfamiliar face matching?. *Cognitive Research: Principles and Implications, 4*, 1-10. https://doi.org/10.1186/s41235-019-0174-3

Bohil, C. J., & Maddox, W. T. (2001). Category discriminability, base-rate, and payoff effects in perceptual categorization. *Perception & Psychophysics, 63*(2), 361–376. https://doi.org/10.3758/BF03194476

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305-327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*(1), 286-291. https://doi.org/10.3758/BRM.42.1.286

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*(8), 1467-1485. https://doi.org/10.1080/17470218.2013.800125

Dale, G., & Arnell, K. M. (2015). Multiple measures of dispositional global/ local bias predict attentional blink magnitude. Psychological Research, 79, 534–547. https://doi.org/10.1007/s00426-014-0591-3

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. https://doi.org/10.3758/BF03193146

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Feng, X., & Burton, A. M. (2019). Identity documents bias face matching. *Perception*, *48*(12), 1163-1174. https://doi.org/10.1177/03010066198778

Feng, X., & Burton, A. M. (2021). Understanding the document bias in face matching. *Quarterly Journal of Experimental Psychology*, *74*(11), 2019-2029. https://doi.org/10.1177/17470218211017902

Fysh, M. C., & Bindemann, M. (2018). The Kent Face Matching Test. *British Journal of Psychology*, *109*(2), 219-231. https://doi.org/10.1111/bjop.12260

Fysh, M. C., & Bindemann, M. (2022). Molistic processing in facial image comparison. *Applied Cognitive Psychology*, *36*(4), 830-841. https://doi.org/10.1002/acp.3975

Gentry, N. W., & Bindemann, M. (2019). Examples improve facial identity comparison. *Journal of Applied Research in Memory and Cognition*, *8*(3), 376. https://doi.org/10.1002/acp.3975

Gong, J., Zhang, Y., Yang, Z., Huang, Y., Feng, J., & Zhang, W. (2013). The framing effect in medical decision-making: a review of the literature. *Psychology, Health & Medicine*, *18*(6), 645-653. https://doi.org/10.1080/13548506.2013.766352

Haarsma, J., Kaltenmaier, A., Fleming, S. M., & Kok, P. (2025). Expectations about presence enhance the influence of content-specific expectations on low-level orientation judgements. Cognition, 254, Article 105961. https:// doi.org/10.1016/j.cognition.2024.105961

Homar, A. R., & Cvelbar, L. K. (2021). The effects of framing on environmental decisions: A systematic literature review. *Ecological Economics*, *183*, 106950. https://doi.org/10.1016/j.ecolecon.2021.106950

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323. https://doi.org/10.1016/j.cognition.2011.08.001

Kantner, J., Vettel, J. M., & Miller, M. B. (2015). Dubious decision evidence and criterion flexibility in recognition memory. *Frontiers in Psychology*, *6*, 1320. https://doi.org/10.3389/fpsyg.2015.01320

Keren, G. (Ed.). (2011). *Perspectives on Framing* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203854167

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23-55. https://doi.org/10.1006/obhd.1998.2781

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123-138. https://doi.org/10.1016/j.jecp.2015.09.030

Lecheler, S., & De Vreese, C. H. (2011). Getting real: The duration of framing effects. *Journal of Communication*, *61*(5), 959-983. https://doi.org/10.1111/j.1460-2466.2011.01580.x

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76(*2), 149-188. https://doi.org/10.1006/obhd.1998.2804

Leynes, P. A. (2021). The effect of test query on recognition event-related potentials (ERPs). *Brain and Cognition*, *155*, 105814. https://doi.org/10.1016/j.bandc.2021.105814

Lindow, S., & Betsch, T. (2018). Child decision-making: On the burden of predecisional information search. *Journal of Cognition and Development*, *19*(2), 137-164. https://doi.org/10.1080/15248372.2018.1436057

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science*, *25*(9), 1663-1673. https://doi.org/10.1177/0956797614541991

Lynn, S. K., Wormwood, J. B., Barrett, L. F., & Quigley, K. S. (2015). Decision making from economic and signal detection perspectives: development of an integrated framework. *Frontiers in Psychology*, *6*, 952. https://doi.org/10.3389/fpsyg.2015.00952

Macmillan, N. A. & Kaplan, H. L. (1985) Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185. https://doi.org/10.1037/0033-2909.98.1.185.

Maddox, W. T., & Bohil, C. J. (2005). Optimal classifier feedback improves cost–benefit but not base-rate decision criterion learning in perceptual categorization. *Memory & Cognition, 33*(2), 303–319. https://doi.org/10.3758/BF03195319

Matthews, C. M., Maurer, D., & Mondloch, C. J. (2022). Recognizing Facial Identity: Prolonged Development During Infancy and Childhood. In K.C. Kadosh (Ed.) *The Oxford Handbook of Developmental Cognitive Neuroscience*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198827474.013.7

Maule, J., & Villejoubert, G. (2007). What lies beneath: Reframing framing effects. *Judgement and Choice: Perspectives on the Work of Daniel Kahneman*, 25-44.

McCaffery, J. M., & Burton, A. M. (2016). Passport checks: Interactions between matching faces and biographical details. *Applied Cognitive Psychology*, *30*(6), 925-933. https://doi.org/10.1002/acp.3281

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. Perception, 44(1), 5–22. https://doi.org/10.1068/p7825

Miller, M. B., & Kantner, J. (2020). Not all people are cut out for strategic criterion shifting. *Current Directions in Psychological Science*, *29*(1), 9-15. https://doi.org/10.1186/s41235-018-0112-9

Mill, R. D., & O'Connor, A. R. (2014). Question format shifts bias away from the emphasised response in tests of recognition memory. *Consciousness and Cognition*, *30,* 91-104. https://doi.org/10.1016/j.concog.2014.09.006

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in Cognitive Sciences*, *22*(9), 794-809. https://doi.org/10.1016/j.tics.2018.06.006

Papesh, M. H., Heisick, L. L., & Warner, K. A. (2018). The persistent low-prevalence effect in unfamiliar face-matching: The roles of feedback and criterion shifting. *Journal of Experimental Psychology: Applied*, *24*(3), 416. https://doi.org/10.1037/xap0000156

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. Quarterly Journal of Experimental Psychology, 70(5), 897–905. https:// doi.org/10.1080/17470218.2015.1136656

Stabile, V. J., Baker, K. A., & Mondloch, C. J. (2024). Criterion shifting in an unfamiliar face-matching task: Effects of base rates, payoffs, and perceptual discriminability. *Journal of Applied Research in Memory and Cognition, 13*(4), 569–581. https://doi.org/10.1037/mac0000157

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149. https://doi.org/10.3758/BF03207704

Stantic, M., Brewer, R., Duchaine, B., Banissy, M. J., Bate, S., Susilo, T., Catmur, C., & Bird, G. (2022). The Oxford Face Matching Test: A non biased test of the full range of individual differences in face perception. Behavior Research Methods, 54(1), 158–173. https://doi.org/10.3758/ s13428-021-01609-2

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089–1104. https://doi.org/10.1037/xge0000109

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, *59*(2), 336–347. https://doi.org/10.1016/j.neuron.2008.05.021

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: Neural and computational mechanisms. *Frontiers in Neuroscience*, *6*, 70. https://doi.org/10.3389/fnins.2012.00070

Summerfield, C., & Egner, T. (2014). Attention and decision-making. In A.C. Nobre & S. Kastner (Eds.), *The Oxford Handbook of Attention* (pp.837-864). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199675111.013.018

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. https://doi.org/10.1037/xap0000108

Towler, A., Kemp, R. I., Burton, A. M., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work?. *PloS one*, *14*(2), e0211037. https://doi.org/10.1371/journal.pone.0211037

Towler, A., Keshwa, M., Ton, B., Kemp, R. I., & White, D. (2021). Diagnostic feature training improves face matching accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(8), 1288. https://doi.org/10.1037/xlm0000972

Tummon, H. M., Allen, J., & Bindemann, M. (2019). Facial identification at a virtual reality airport. *i-Perception*, *10*(4), 2041669519863077. https://doi.org/10.1177/2041669519863077

Tyler, R., Towler, A., Kemp, R. I., & White, D. (2022). Let's talk about faces: Identifying faces from verbal descriptions. *British Journal of Psychology*, *114*(1), 262-281. https://doi.org/10.1111/bjop.12610

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458. https://doi.org/10.1126/science.7455683

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161-204. https://doi.org/10.1080/14640749108400966

White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd effects in unfamiliar face matching. Applied Cognitive Psychology, 27(6), 769 777. https://doi.org/10.1002/acp.2971

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS one*, *9*(8), e103510. https://doi.org/10.1371/journal.pone.0103510

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. https://doi.org/10.1037/xap0000114