The Efficiency of Demography in Face Perception

Clara Colombatto[1], Stefan Uddenberg[2], & Brian J. Scholl[1]

[1]Yale University, [2] Princeton University

Running Head : Facial Demography

Address for correspondence : Brian Scholl
Department of Psychology
Yale University
Box 208205
New Haven, CT 06520-8205

Email : brian.scholl@yale.edu
Phone : 203-432-4629

Word Count : 7320 (Main text + footnotes)
Version : 7/6/21 — In press, *Attention, Perception, & Psychophysics*

Abstract (200 words)

When we look at a face, we cannot help but 'read' it: beyond simply processing its identity, we also form robust impressions of both transient psychological states (e.g. surprise) and stable character traits (e.g. trustworthiness). But perhaps the most fundamental traits we extract from faces are their social demographics — e.g. race, age, and gender. How much exposure is required to extract such properties? Curiously, despite extensive work on the temporal efficiency of extracting both higher-level social properties (such as competence and dominance) and more basic characteristics (such as identity and familiarity), this question remains largely unexplored for demography. We correlated observers' percepts of the race/age/gender of unfamiliar faces viewed at several brief durations (and then masked) with their judgments after unlimited exposure. Performance reached asymptote by 100 ms, was above chance by only 33.33 ms, and had a similar temporal profile to detecting faces in the first place. This was true even when the property to be reported wasn't revealed until after the face had disappeared, and when the faces were matched for several lower-level visual properties. Collectively, these results demonstrate that the extraction of demographic features from faces is highly efficient, and can truly be done at a glance.

Keywords

Face perception; Race perception; Age perception; Gender perception; First impressions

Statement of Significance (98 words)

When looking at someone's face, you quickly form a wide variety of impressions, e.g. of trustworthiness or extraversion, "at a glance": people are just as good at perceiving such properties from quickly flashed faces as they are during extended viewing. But we also readily have demographic impressions — e.g. race, age, and gender. Here we show that the formation of such demographic impressions is, if anything, even more efficient: not only are they formed at an (even shorter) glance, but we see them as soon as we can see that there is a face in the first place.

If any old picture is worth a thousand words, then pictures of *faces* may be worth considerably more than that: when viewing faces, we form impressions of a wide variety of properties, even beyond the extraction of identity information (for a review see Todorov, 2017). Some of these properties reflect relatively stable traits — such as how trustworthy, competent, or dominant a person seems to be. Other properties reflect relatively transient states — as when a person looks momentarily focused, distracted, or disappointed. And still other properties reflect our personal experience or preferences — as when faces strike us as familiar, or as attractive. The fact that we so readily ‘read’ faces in this way seems extraordinary given that all faces share the same basic features and overall configuration, and given that so many of our impressions of faces seem so ineffable.

Perceiving Demography?

Some of the most intriguing traits that we discern from faces, however, go beyond those already mentioned: we also readily perceive *demographic* characteristics, such as race, age, and gender. Such demographic features are undeniably important, given how they influence both other psychological processes (such as memory; e.g. Meissner & Brigham, 2001) and real-world outcomes (as when a person’s perceived race influences police officers’ split-second decisions about whether to shoot a potential perpetrator; e.g. Correll et al., 2007). But there is some disagreement about just how fundamental demographic properties are in social perception.

In some ways, the perception of properties such as race, age, and gender seem even more foundational than other features. For example, whereas it is controversial whether facial expressions of emotion are universal (e.g. Gendron et al., 2014), every society has both men and women.[1] And whereas there is considerable debate about the degree to which split-second inferences about social traits like trustworthiness are reliable signals (e.g. Todorov et al., 2015), characteristics such as perceived age are especially honest cues. (Indeed, aging is associated with a number of structural and textural changes in faces [e.g. Berry & McArthur, 1986; George

[1] Note that we use the term “gender” throughout this manuscript to refer to the perception of sex.

& Hole, 2000] that are notoriously difficult to conceal, though not for any lack of trying by the cosmetics industry.) And more generally, perceived demographic traits often lead us to categorize others on the basis of social groups (such as race; e.g. Cloutier et al., 2005), which does not happen so readily on the basis of social traits such as extraversion.

At the same time, there are other indications that at least some demographic traits may not be so foundational in social perception after all. For example, perceived race and perceived gender seem to radically diverge in terms of how irresistibly they are encoded into memory (Cosmides et al., 2003). In “who said what?” memory tests, for example, errors rarely if ever cross gender boundaries regardless of whether gender is made salient or not. But for perceived race the situation is more complicated: while early studies suggested that race is similarly automatically encoded, later studies revealed that such memory traces largely vanish when there are other competing ‘coalitional’ cues — e.g. involving shirt color (Kurzban et al., 2001; Pietraszewski et al., 2014) or political affiliation (Pietraszewski et al., 2015). These results suggest that the perception of demographic cues such as race may not be so primitive after all — such that perceived ‘race can be erased’ (while perceived gender cannot).

The Current Studies: Seeing, Fast and Slow

The current experiments seek to contribute to the larger project of determining how foundational demographic properties are in social perception, using an especially crude but apt metric: *temporal efficiency*. Studies of the timecourse of face perception are legion. And one lesson from past work is that face perception, in general, is *fast*. For example, past work has identified the minimal response time it takes to detect the presence of a human face among distractors (~ 240-290 ms; Rousselet et al., 2003), to detect the presence of a familiar face (~ 360-470 ms; Barragan-Jason et al., 2012; Besson et al., 2012), or to recognize the identity of a particular individual (~ 260 ms; Besson et al., 2017). And beyond these brute speed measurements, face perception also seems highly *efficient:* Previous work suggests that holistic processing of faces relies on coarse yet rapidly available low spatial frequencies (Goffaux &

Rossion, 2006), and can occur after only ~50 ms of exposure time (Richler et al., 2009; 2011).[2] And similarly, a tidal wave of prominent work in social perception has shown how much exposure to a face is required in order to form impressions of properties such as trustworthiness (Willis & Todorov, 2006), threat (Bar et al., 2006), and extraversion (Borkenau et al., 2009). For example, Willis and Todorov (2006) showed observers faces of differing levels of trustworthiness (and other properties such as competence and aggressiveness). Each face was viewed twice — once for a limited duration, after which it was masked, and once for an unlimited time. The key analyses then involved correlations between speeded and unspeeded trait judgments, and the key result was that such correlations tended to reach asymptote by approximately 100 ms — such that no further information about such traits was extracted after that point.

Curiously, despite all of this work on the timecourse of face processing for higher-level social properties, there has been almost no exploration of the efficiency with which basic demographic properties are extracted. The current experiments aim to fill this gap, in what is to our knowledge the first behavioral exploration of the amount of exposure to faces that is required in order to perceive demography. Following the logic of Willis and Todorov (2006) as reviewed above, we correlated observers' percepts of the race/age/gender of unfamiliar faces viewed at several brief durations (and then masked) with their unspeeded judgments of the same (unmasked) faces.

In Experiment 1, each observer had to indicate only a single demographic property (either perceived race, age, or gender) throughout their entire session. In Experiment 2, we explored whether the temporal profile of demographic perception changed substantially when

[2] We limit ourselves here to measures of the speed of behavioral performance, but of course there are also many studies of the speed with which various kinds of neural signatures can be detected during face perception. In particular, an extensive body of work has explored the timecourse of face processing using scalp and intracranial recordings of neural activity (e.g. Bentin et al., 1996). This work has encompassed many different perceived facial features, including demography (e.g. Colombatto & McCarthy, 2017) — and indeed recent work with magnetencephalography has shown that such signatures emerge even earlier for perceived race and perceived gender than for identity and familiarity (Dobs et al., 2019).

the property to be reported wasn't revealed until after the face had disappeared, and observers thus had to extract and hold in memory multiple properties at once. And in Experiment 3, we asked whether perceiving demography follows a similar timecourse to the detection of the presence of a face in the first place (and this experiment also empirically validated the effectiveness of our face masks). In Experiments 2 and 3, we also used faces that were tightly cropped and matched for mean luminance, thus controlling for several lower-level visual properties.

Experiment 1: Demography in the Blink of an Eye

In an initial exploration of how much exposure is required in order to extract demographic information from faces, observers viewed the same faces (which varied in their race, age, and/or gender, as depicted in Figure 1a) either with unlimited exposure time and without any masking, or when the faces were masked (as depicted in Figure 1b) after a delay ranging from 16.66 ms to 1000 ms.

Method

Subjects

Twenty-four members of the Yale/New Haven community (8 observers for each of the 3 tasks; 10 females; average age=19.17 years, *SD*=0.96 years; 15 self-identified as White; 5 Asian; 2 Black; 1 Latino; 1 both Asian and White) participated in exchange for course credit. This preregistered sample size was chosen before data collection began based on pilot data, and was fixed to be identical in all 3 experiments reported here.[3] All experimental methods and

[3] Our key conclusions from this study stem not from any particular statistical comparison (the power for which could be explicitly computed), but rather from the overall pattern of absolute correlations at differing exposure durations — and in particular from the fact that correlations in the 50ms-exposure conditions were extremely high. Put in graphical terms, our conclusions are based on the salient inverted-L shaped patterns depicted in Figure 3, in which the correlation in the 50ms-exposure condition was near ceiling. That said, this sample size was determined based on a pilot study identical in design to Experiment 1, with 24 observers for each of the race, age, and gender tasks, with this number originally chosen to match published work using the same method (Willis & Todorov, 2006). In these pilot data, a

procedures were approved by the Yale University Institutional Review Board, and were preregistered (see https://osf.io/pzrhw/?view_only=c85385291bf04d278e5e6609034b17ec).

Apparatus

Stimuli were presented on a DELL M992 CRT monitor with a 60Hz refresh rate, using custom software written in Python with the PsychoPy libraries (Peirce, 2019). Observers sat in a dimly lit room without restraint approximately 60cm from the display, which subtended 33.57°×25.49°; all visual extents reported below were computed based on this viewing distance.

Stimuli

The kinds of faces and masks used in this experiment are depicted in Figures 1a and 1b. The faces were chosen for each task separately, with the goals of (1) maximizing *discriminability* along the dimension of interest, while (2) maximizing *variance* in the other dimensions (to the degree allowed by the face databases themselves). For the age task, 60 young faces (18-28 years old) and 60 old faces (66-88 years old) were drawn from the UT Dallas Park lab Face Database (Minear & Park, 2004). These faces were all White, and there were an equal number of males and females in each group (30 each). (See the left column of Figure 1a for sample faces, which are similar but not identical to those used in the experiment. For full stimuli, see http://agingmind.utdallas.edu/download-stimuli/face-database/). For the race task, 60 White and 60 Black faces were drawn from the Chicago Face Database (Ma et al., 2015). These faces were selected such that they would have the greatest discriminability along the Black-White race dimension, based on norming data from a large sample of participants (N=1087) who categorized each face as either Asian, Black, Hispanic/Latino, White, or Other, in a self-paced survey (for details, see Ma et al., 2015). Discriminability was operationalized as the difference in proportions of participants who rated a face as Black vs. White (for faces of models who self-identified in these ways). This was done so that we could compare observers' responses to their

sample size of 24 proved sufficient (a) to obtain 90% power to detect an effect size of 4.39, when calculating whether performance was above chance at the key 50ms exposure duration, and (b) to precisely estimate the correlation at the key 50ms exposure duration with a standard deviation across observers of only 0.08 (averaged across the correlations for perceived race, age, and gender).

responses in other conditions, as well as to the 'ground truth' — defined as the model's self-reported race and confirmed by a separate group of participants.[4] We selected the 60 most discriminable faces for each race, and there were an equal number of males and females in each group (30 each; see the middle column of Figure 1a). For the gender task, 60 male and 60 female faces were drawn from the Chicago Face Database (Ma et al., 2015). These faces were selected such that they would have the greatest discriminability along the male-female dimension, again using gender categorizations from a separate sample of participants (for details, see Ma et al., 2015), such that the 'ground truth' corresponds to the model's self-reported gender as confirmed by a separate group of participants. There were an equal number of White, Black, Asian, and Latino faces in each group (15 each; see the middle column of Figure 1a). Some individual faces were used in both the race and gender tasks, but (given that task was a between-subjects factor), each observer only made a single speeded judgment on any given face.

A different set of 120 masks was created for each task, as follows. First, each face image was subdivided into 0.68° square regions, each of which was categorized as either blank or colored (based on whether the average RGB values in that region were above some threshold, which differed by task). Each mask was then created by first choosing a random base image from the same task. Each blank region was then left unchanged, while each colored region was replaced by a different randomly chosen colored region (randomly rotated by either 0°, 90°, 180°, 270°) from a randomly chosen image from that same task. As depicted in Figure 1b, this gave rise to mask images that had recognizable head-shaped silhouettes, but without any

---

[4] By 'ground truth,' we do not mean to imply that others' impressions of race or gender must correspond to some underlying physical truth, since of course this is demonstrably false for many related impressions. Many first impressions are influenced in part by morphological features (e.g. width-to-height ratio in the case of perceived dominance), but are also highly susceptible to perceivers' characteristics (e.g. their race, gender, group membership, and stereotype associations; Xie et al., 2019, in press). And demography categorization is no exception, since it can be influenced by perceivers' characteristics (such as interracial exposure; Freeman et al., 2016) — although impressions of demography also seem more consistent across observers (e.g. Hehman et al., 2017). To account for these sources of variability, performance is quantified in the current work as consistency both within and across observers, with 'ground truth' referring to the latter.

recognizable facial features (demographic or otherwise) — thus controlling for many of the low-level properties of the images. The resulting set of masks for each task were then randomly assigned to each trial, differently for each observer. (And the efficacy of these masks is empirically validated in Experiment 3.)

Procedure

The procedures employed in this experiment are summarized in Figure 2. Each trial began with a central black fixation cross (0.78°×0.78°) on a white background for 500 ms, followed by a centered face image (22.75°×16.10° in the race and gender tasks, 21.35°×16.10° in the age task). On Unspeeded blocks, the face was surrounded by a green frame (with a stroke of 0.17°) and remained visible until response. On Speeded blocks, the face was surrounded by a red frame (with the same stroke width) and remained visible for a limited exposure time, after which it was immediately replaced by a mask (surrounded by the same frame) for 500 ms, followed by a blank screen (with a green frame) until response.

Observers pressed a key from 1 to 5 to indicate their response, where 1 and 5 represented the extremes ("Definitely White" and "Definitely Black" in the race task, "Definitely Young" and "Definitely Old" in the age task, and "Definitely Female" and "Definitely Male" in the gender task), and 3 represented "Not sure". These labels were present just below the five numerals themselves (all drawn in 25-point Helvetica) throughout each trial, with the numerals vertically separated from the image border by 1.18° and from the labels by 0.98°. (The numeral '1' was presented with its leftmost point 10.30° to the left of the display center; the numeral '5' was presented with its rightmost point 10.33° to the right of the display center, and the other three numerals were evenly spaced in between 1 and 5. Each of the three labels — Definitely X, Not Sure, Definitely Y — was roughly centered below its relevant numeral.)

Design

Observers completed ten 24-trial-long blocks. Half were Unspeeded, and half were Speeded — with one block for each of five exposure durations (16.66, 50, 100, 500, and 1000 ms). The order of blocks was randomized (separately for each observer), and there was a self-paced rest period after each block. The experiment began with 6 practice trials (one Unspeeded, and one for each of the 5 exposure durations), the results of which were not recorded. Each of the 120 faces was then presented twice over the course of the experiment — once during an Unspeeded block, and once during a randomly chosen Speeded block. The 24 faces within each block were fully counterbalanced in terms of the relevant demographic variables.

Results and Discussion

The mean Pearson correlations between observers' Speeded and Unspeeded responses for each exposure duration are depicted in Figure 3a, separately for perceived race, age, and gender. Inspection of this figure suggests a simple response pattern that held for each task: correlations were relatively low for 16.66 ms, but thereafter (by 50 ms) were at ceiling. The striking nature of this pattern makes it unclear what more could be gained by statistical analyses, but we nevertheless quantified these effects as follows. A series of *t* tests confirmed that the average correlations were above chance for all three perceived features at all exposure durations ($\alpha_{corrected}$=.01, all *r*s≥.82, *n*s=24 faces in each task/exposure duration, *t*s[22]≥6.82, *p*s<.001, *d*s≥2.91) except for 16.66 ms (perceived race: *r*=.05, *n*=24, t[22]=0.24, *p*=.813, *d*=0.10; perceived age: *r*=-.11, *n*=24, *t*[22]=0.52, *p*=.607, *d*=0.22; perceived gender: *r*=.31, *n*=24, *t*[22]=1.53, *p*=.139, *d*=0.65). Additional within-subjects *t* tests confirmed that the correlations for 16.66 ms trials differed from all other exposure durations in that same task ($\alpha_{corrected}$=.005, all *t*s≥6.11, *p*s≤.002, *d*s≥4.99), but that no other exposure duration differed from any of the others (all *t*s[7]≤3.79, *p*s≥.007, *d*s≤2.87), except for 50 vs. 1000 ms in the gender task (*t*[7]=4.32, *p*=.003, *d*=3.27). (Note that some comparisons have fewer degrees of freedom, because of differences in how frequently some observers in each condition provided entirely uniform responses.)[5]

[5] Response times are not analyzed here (as per the pre-registration) (a) because observers were not asked to make speeded responses, and (b) because the speed at which observers respond does not directly bear

When Speeded responses were correlated not with Unspeeded responses but rather with the ground truth (i.e. the actual demographics of the people in the images), these same patterns all held, as depicted in Figure 3b (analyses not included).[6]

The stark patterns of responses in Figure 3a suggest that observers were guessing at 16.66 ms, but were certain at the other exposure durations, and this is further supported (again for each of the three tasks) by the distributions of particular responses at each exposure duration, as depicted in Figure 4. In particular, this figure makes it clear that observers used the extremes of the scale (“Definitely”) and almost nothing else at all exposure durations except 16.66 ms, where the vast majority of their responses were “Not sure”. This impression was verified via a series of Kolmogorov-Smirnov Tests comparing the distributions of responses at each exposure time: the distribution of responses at the shortest exposure time (16.66ms) differed substantially from all other exposure times in all conditions (all Ds$\geq$.38, all ps<.001), and no other test reached significance (all Ds<0.04, ps>.919).

The results of this experiment collectively support an especially clear-cut conclusion: perceived race, age, and gender are reliably extracted from faces even with only 50 ms of exposure, with no further benefit beyond this point.

Experiment 2: 1 vs. 2 Demographic Features

Are a face’s demographic features available to perceivers only when the feature in question is explicitly highlighted in advance of seeing the face — as was the case for Experiment 1 (where the task was always the same for each trial for a given observer)? Although we do not focus in depth on this question in the current project, we do provide the most minimal possible test, by asking whether observers are able to extract *two* demographic properties at once with no

---

on the particular types of questions about exposure that we are asking. (Nevertheless, full response data is available in the raw data file accompanying our submission.)

[6] This pattern of results (and all subsequent ones reported in this manuscript) was unaltered when the raw Pearson correlations were converted to Fisher’s Z scores prior to computing their averages.

cost— when they don't know which one will be queried until after the faces disappear. We simply highlighted the target demographic property before the face was displayed on half of trials, while on the other half of trials observers were only asked (about the face's race or gender) after the face was no longer visible.

This experiment also featured a new set of more controlled stimuli, to help ensure that observers relied on the perception of demographic information per se, rather than on specific lower-level cues that might be correlated with a given feature. In particular, faces were tightly cropped (to eliminate cues such as hair length that might otherwise be used as proxies in the gender task) and were matched for mean luminance (to eliminate brute differences in skin tone that might otherwise be used as a proxy in the race task; see Figure 1 for sample stimuli). (The age task was simply eliminated altogether. Whereas the most salient image characteristics for the race and gender tasks — viz. lightness and hair length — seem relatively extrinsic, the perception of age relies on more subtle and numerous intrinsic features, including head shape, facial proportions, skin texture, wrinkles, etc.; Mark et al., 1980; O'Toole et al., 1997. And while a cropped female face still looks female — and a luminance-matched Black face still looks Black — a wrinkle-free old face no longer looks old.)[7] And for added precision, we also replaced the extraneous 1000 ms exposure duration with an intermediate duration of 33.33 ms.

Method

This experiment was identical to Experiment 1, except as noted here. Eight members of the Yale/New Haven community (4 females; average age=19.25 years, *SD*=0.71 years; 4 self-identified as White; 3 Asian; 1 both Asian and White) participated in exchange for course credit, with this preregistered sample size chosen to exactly match that of Experiment 1 (see https://osf.io/pzrhw/?view_only=c85385291bf04d278e5e6609034b17ec).

---

[7] Another manipulation that is commonly used to assess the relative contribution of low-level features to face processing is face *inversion* — which attempts to hold facial features constant while disrupting face-specific configural processing. Our approach of controlling for specific extrinsic features seemed preferable to this approach for the current study, however, given that many researchers argue that face-inversion effects are merely quantitative rather than qualitative, and that they can be less than compelling in practice (e.g. Richler et al., 2011; Rossion & Gauthier, 2002; Valentine, 1988).

A set of 120 faces (30 Asian Females, 30 Asian Males, 30 White Females, 30 White Males) were selected from the Chicago Face Database (Ma et al., 2015). Each face was placed into an approximately oval mask using Psychomorph (Tiddeman et al., 2005) to render the hairline and ears invisible. As depicted in Figure 1a, the resulting stimuli were converted to grayscale, and matched for mean luminance using the SHINE toolbox in MATLAB (Willenbockel et al., 2010). And a new set of 600 masks was generated from these faces via the same procedure as in Experiment 1 (with an example depicted in Figure 1b).

The procedures employed in this experiment are summarized in Figure 2. On Prompt Before blocks, the prompt "Race?" (2.36°×0.68°) or "Gender?" (3.21°×0.68°) appeared (in 25-point Helvetica) for 500 ms in the center of the display immediately before the face appeared, while on Prompt After blocks it appeared for the same duration immediately after the mask's presentation. Consequently, the response scale labels were shown on each trial only on the empty response screen(s) following the faces (i.e. following the mask on Prompt Before blocks, and following the prompt on Prompt After blocks).

Observers completed twelve 60-trial-long blocks (in a different random order for each observer) — half Prompt Before and half Prompt After. For each prompt type, one block was unspeeded, while the remaining five blocks had exposure durations of 16.66, 33.33, 50, 100, and 500 ms. The experiment began with 6 practice trials of each type (one Unspeeded, and one for each of the 5 exposure durations), the results of which were not recorded. Each of the 120 faces was presented on either Prompt Before or Prompt After blocks, at each of the six exposure durations, for a total of 720 trials per observer.

<u>Results</u>

The mean correlations between observers' Speeded and Unspeeded responses for each exposure duration and each prompt condition are depicted in Figure 5, separately for perceived race and perceived gender. Inspection of this figure suggests three patterns (all of which were apparent for both perceived race and perceived gender): (a) correlations reached asymptote (with no further advantage for additional exposure) by 100 ms; (b) correlations were already above chance — but below asymptote — at both 33.33 and 50 ms; and (c) the before-vs.-after prompt manipulation had almost no discernible effect.

These impressions were verified by the following statistical analyses. A series of *t* tests confirmed that average correlations between observers' Speeded and Unspeeded responses were above chance for both tasks at all exposure times ($\alpha_{corrected}$=.01, all *r*s≥.40, *n*s=60 faces in each task/exposure duration, *t*s[58]≥3.36, *p*s≤.001, *d*s≥0.88) except for 16.66 ms for both perceived race (*r*=.08, *n*=60, *t*[58]=0.60, *p*=.550, *d*=0.16) and perceived gender (*r*=.03, *n*=60, *t*[58]=0.25, *p*=.805, *d*=0.07), and also 33.33 ms for perceived gender (*r*=.27, *n*=60, *t*[58]=2.14, *p*=.037, *d*=0.56). Additional within-subjects *t* tests confirmed that the correlations at the shortest exposure duration (16.66 ms) differed substantially from both 100 and 500 ms in both tasks ($\alpha_{corrected}$=.005, all *t*s[6]≥5.95, *p*s≤.001, *d*s≥4.86); that the 33.33 ms exposure time differed substantially from 500 ms in both tasks (all *t*s[7]≥4.94, *p*s≤.002, *d*s≥3.73), and from 100 ms in the gender task (*t*[7]=8.22, *p*s<.001, *d*=6.21); and that the 50 ms exposure time differed substantially from both 100 and 500 ms in the gender task (all *t*s[7]≥4.51, *p*s=.003, *d*s≥3.41). No other exposure duration differed from any of the others (all *t*s≤4.00, *p*s≥.005, *d*s≤3.03). (Note that some comparisons have again fewer degrees of freedom, because of differences in how frequently some observers in each condition provided entirely uniform responses.)

Perhaps most critically, given the purpose of this experiment, there was no overall difference in performance between Prompt Before and Prompt After trials, for either the gender task or the race task — and this was true both overall (i.e. for within-subject comparisons at each limited exposure duration; perceived race: *t*[37]=0.91, *p*=.367, *d*=0.30; perceived gender: *t*[37]=0.89, *p*=.378, *d*=0.29) and for each individual exposure duration ($\alpha_{corrected}$=.01, all *t*s≤3.18, *p*s≥.025, *d*s≤2.84). (And when Speeded responses were correlated not with Unspeeded responses but rather with the ground truth, these same patterns all held — though we do not report these analyses. The overall patterns of results in each task were also unaltered by the other perceived demographic trait. (The timecourse of gender perception was similar for faces perceived as White or Asian, and [vice versa] the timecourse of race perception was similar for faces perceived as Male or Female — both overall and in the [null] effect of prompt timing.)

These results are further supported by the distributions of particular responses at each exposure duration, as depicted in Supplementary Figure 1. In particular, this figure makes it clear that observers responded “Not Sure” on the vast majority of trials with a 16.66 ms exposure duration, while they used the whole scale at 33.33 ms, and mostly just the extremes of the scale (“Definitely”) at longer exposure durations, with no discernible differences between Prompt Before and Prompt After trials. These impressions were verified via a series of Kolmogorov-Smirnov Tests comparing the distributions of responses at each exposure time: the distribution of responses at the shortest exposure time (16.66ms) differed substantially from all other exposure times in all conditions (all Ds$\geq$.20, all ps<.001), and no other test reached significance (all Ds<0.04, ps>.919). There was also no difference in the distributions of responses between Prompt Before and Prompt After trials (all Ds<0.09, ps>.317, with a marginal effect at 500 ms in the race task, D=.12, p=.060).

Discussion

The core result of this experiment was that observers were no worse at extracting race and gender in faces when they did not know which one would be relevant until the face had disappeared; in other words, there was no cost to extracting two different demographic variables compared to only one. These conclusions remain preliminary, as they are based on a null effect. And in addition, we note that this experiment was not designed to investigate comparisons at particular timepoints, but rather only a general pattern of results across varying exposure times. Nevertheless, there was no hint of a difference in this key comparison — providing preliminary evidence for the equivalence of extracting 1 vs. 2 features.

In both the gender and race tasks, performance again reached asymptote surprisingly early, though these results were weaker in two ways compared to those of Experiment 1. First, whereas performance reached asymptote by 50 ms in Experiment 1, here this was not true until 100 ms of exposure. Second, the asymptote itself was notably lower — roughly 75% instead of nearly perfect. These results suggest that the cropping and luminance-matching employed in this experiment (and in Experiment 3) were important after all, since the designs were otherwise similar: apparently (if unsurprisingly) observers do make use of simple proxies (such as skin tone and hair length) when they are available, and eliminating these possibilities both impairs performance and requires more online exposure to the faces. (Of course, some of these differences could also reflect the change in the specific races that were contrasted in this experiment.) At the same time, however, it was still especially notable that even under such controlled conditions, performance was still reliably above chance even at 33.33 ms (and that performance still reached asymptote by 100 ms).

Experiment 3: Demography vs. Detection

The efficiency of demography in face perception, as measured in Experiments 1 and 2, is partly impressive in an absolute sense (since durations of only 50-100 ms are extremely brief for any type of social perception), but they are also impressive in a relative sense — since these durations are comparable to the efficiency of face detection in the first place (e.g. Crouzet et al., 2010). This raises the provocative possibility that perhaps perceiving demography is best characterized not as some extra stage of face perception, but rather as an irresistible process triggered automatically by the detection of faces. In short, perceiving demography might be part of what it means to perceive faces in the first place. Accordingly, in this experiment we asked directly whether perceiving demography requires any additional processing time past face detection itself: as soon as you see a face, do you also see its race/age/gender (cf. Grill-Spector &

Kanwisher, 2005)? Observers viewed scrambled faces that were preceded briefly by either an unscrambled face (on half of trials) or a different scrambled face (on the other half of trials), and for each trial they reported both (a) whether an intact face was present ("Present"/"Absent"; Grill-Spector & Kanwisher, 2005; Thorpe et al., 1996), and (b) its demographic properties ("White"/"Asian" in the race task, or "Female"/"Male" in the gender task). (We also included both Masked and Unmasked blocks, in order to further demonstrate the effectiveness of the masks. And we only tested 16.66, 33.33, and 50 ms, since performance had already reached asymptote by the longer exposure durations in the previous experiments — and since the unlimited-exposure trials were always redundant with the ground truth itself.)

Method

This experiment was identical to the Prompt Before blocks in Experiment 2, except as noted here. Eight members of the Yale/New Haven community (5 females; average age=20.38 years, *SD*=1.85 years; 3 self-identified as White; 2 Asian; 1 Black; 1 Hispanic/Mexican; 1 both Asian and White) participated in exchange for course credit, with this preregistered sample size chosen to exactly match that of Experiments 1 and 2 (see https://osf.io/pzrhw/?view_only=c85385291bf04d278e5e6609034b17ec).

A new set of 360 masks was generated from the faces according to the procedure described in Experiment 1 (with an example depicted in third column of Figure 1b). And an additional 120 stimuli were generated using the same procedure, to serve as scrambled faces in the detection task.

The procedures employed in this experiment are summarized in Figure 2. On Face Present trials, the prompt was followed by a face stimulus presented for a limited duration, and on Face Absent trials, the prompt was instead followed by a scrambled face. On Masked blocks, the first stimulus (either a scrambled or unscrambled face) was replaced by a mask (of the same size, and also surrounded by a red frame) appearing for 500 ms (as in Experiments 1 and 2). On

Unmasked blocks, the first stimulus was instead replaced by a blank screen for the same duration.

After the stimuli disappeared, observers were first prompted to press one of two keys to indicate whether they had perceived a face at all (with the reminders “Present” or “Absent” appearing — in the same font, size, and vertical location — as the numerical scale from Experiment 2), with the relevant label turning red (#010101) for 100 ms upon response. After a 100 ms delay, observers then completed the demography task, reporting either the race or gender of the face as in Experiment 2 (now with binary choices, with the “Present”/”Absent” labels replaced with either “White”/”Asian” or “Female”/”Male”).

Observers completed six 120-trial-long blocks (in a different random order for each observer) — half Masked and half Unmasked. For each mask type, three blocks were presented (16.66, 33.33, 50.00 ms). The experiment began with 5 Masked and 5 Unmasked practice trials (one for each of the 3 exposure times, along with two 500 ms exposure times to help introduce observers to the task), the results of which were not recorded. The faces presented in Masked vs. Unmasked blocks were always distinct, and were randomly chosen separately for each observer.

Results

The mean correlations between observers’ race/gender responses and the ground truth are depicted in Figure 6a, separately for Masked vs. Unmasked blocks, and for each of the three exposure durations. And the mean sensitivity in the detection task (expressed as d’) is similarly depicted for these different types of trials in Figure 6b. Inspection of these figures suggests three patterns (each of which was apparent for both perceived race and perceived gender): (a) correlations for Masked blocks were above chance at both 33.33 and 50 ms, with the correlations at 50 ms roughly equivalent to those observed for the longest exposure durations in Experiment 2; (b) the masks, as expected, were highly effective — impairing performance by 20-60%; and (c) performance in the detection task (now measuring sensitivity) followed a similar relative pattern.

These impressions were verified by the following statistical analyses. A series of *t* tests confirmed that average correlations between observers' demography responses and the ground truth were above chance for both tasks at all exposure times ($\alpha_{corrected}$=.008, all *r*s≥.48, *n*s=30 faces in each task/exposure duration/masking condition, *t*s[28]≥2.89, *p*s≤.007, *d*s≥1.09) except in Masked blocks for 16.66 ms for both perceived race (*r*=.09, *n*=30, *t*[28]=0.46, *p*=.647, *d*=0.18) and perceived gender (*r*=-.09, *n*=30, *t*[28]=0.45, *p*=.654, *d*=0.17), and also 33.33 ms for perceived gender (*r*=.31, *n*=30, *t*[28]=1.73, *p*=.094, *d*=0.66). Additional within-subjects *t* tests confirmed that these correlations generally did not differ from each other ($\alpha_{corrected}$=.008, all *t*s≤6.75, *p*s≥.017, *d*s≤9.54), with three exceptions: 16.66 vs. 50 ms for Masked perceived gender (*t*[3]=6.75, *p*=.007, *d*=7.80); 16.66 vs. 50 ms for Unmasked perceived gender (*t*[7]=3.89, *p*=.006, *d*=2.94); and 33.33 vs. 50 ms for Masked perceived race (*t*[6]=5.69, *p*=.001, *d*=4.65). (Note that some comparisons again have fewer degrees of freedom, because of differences in how frequently some observers in each condition provided entirely uniform responses.)

Most importantly, given the purpose of this experiment, a series of *t* tests confirmed that observers' relative sensitivities in the detection task were analogous to the relative patterns of correlations in the demography task (although the two tasks cannot be directly compared given the different performance measures — with one using sensitivity and the other using correlations). In particular, sensitivity was above chance for both tasks at all exposure times ($\alpha_{corrected}$=.008, all *d'*≥2.42, *t*s[7]≥5.18, *p*s≤.001, *d*s≥3.92) except for 16.66 ms in Masked blocks for both perceived race (*d'*=0.03, *t*[7]=0.52, *p*=.619, *d*=0.39) and perceived gender (*d'*=0.06, *t*[7]=0.84, *p*=.430, *d*=0.63) — as is clear from the two lowest points in Figure 6b. Additional within-subjects *t* tests confirmed that in Masked blocks, the sensitivities at 33.33 and 50 ms differed substantially from 16.66 ms ($\alpha_{corrected}$=.008, all *t*s[7]≥4.79, *p*s≤.002, *d*s≥3.62) and from each other in the gender task (*t*[7]=3.98, *p*=.005, *d*=3.01) — with no other exposure duration differing from any of the others (all *t*s[7]≤3.09, *p*s≥.017, *d*s≤2.34).

Finally, as expected, performance was generally worse (and often much, much worse) on Masked vs. Unmasked blocks (for both race and gender tasks). For the demography tasks, this was true both overall (i.e. for within-subject comparisons at each limited exposure duration; perceived race: $t[17]=6.26$, $p<.001$, $d=3.04$; perceived gender: $t[19]=8.35$, $p<.001$, $d=3.83$) and for each individual exposure duration ($\alpha_{corrected}=.017$, all ts≥3.43, ps≤.017, ds≥2.59). For the detection task, this was also true both overall (i.e. for within-subject comparisons at each limited exposure duration; perceived race: $t[23]=6.01$, $p<.001$, $d=2.50$; perceived gender: $t[23]=6.56$, $p<.001$, $d=2.74$) and for each individual exposure duration ($\alpha_{corrected}=.017$, all ts[7]≥3.13, ps≤.017, ds≥2.36), except for 50 ms in the race task ($t[7]=2.22$, $p=.062$, $d=1.68$).

These patterns of results are further supported by the distributions of particular responses for Face Present trials at each exposure duration, as depicted in Supplementary Figure 2. In particular, this figure makes it clear that observers detect the face and used both response options for demography in all cases except 16.66 ms, where the vast majority of their responses were “Absent”. And importantly, this pattern vanished for unmasked faces, where observers were able to detect the presence of the face even at 16.66 ms.

Discussion

Beyond empirically demonstrating the effectiveness of the masks, the key result of this experiment was that the exposure duration at which observers were first able to reliably perceive the race and gender of a face (33.33 ms) was the same exposure duration at which they were first able to detect the presence of a face at all. This is consistent with the possibility that perceiving demography is not an “extra step” in face perception, but is rather triggered automatically upon face detection, such that as soon as you see a face you also extract its demographic properties. In this way, the perception of demographic properties might be an intrinsic part of seeing a face in the first place.

General Discussion

How much exposure is required in order to extract socially relevant information from faces? This question has been extensively studied in the context of many dimensions — from identity and familiarity to competence and trustworthiness. But this question hasn't been previously asked in the context of the demographic properties of faces, and so the current project was designed to fill this curious gap. The answer, it turns out, is that the perception of demography from faces is efficient, indeed. Observers' ability to extract properties such as race, age, and gender from faces reached asymptote after only 100 ms of exposure, and was above chance by only 33.33 ms. In principle, this ability could reflect a substantive role for the processing of demography in face perception, or it could merely reflect superficial responses based on how light or dark the image as a whole is (as a crude proxy for perceived race) or whether there is any stimulus at all in the region just above the shoulders (as a crude proxy for hair length and thus perceived gender). We controlled for these superficial possibilities in Experiments 2 and 3 by tightly cropping the images and by matching the images for mean luminance. The results clearly indicated that such proxies were used (since performance as a whole declined), but they did not affect the conclusions as stated above: even when such superficial strategies were thwarted, the perception of demography from faces still reached asymptote after only 100 ms of exposure, and was still above chance by 33.33 ms.

Beyond these central results, there were also two additional hints that the perception of demography from faces is especially efficient. First, performance didn't differ (in Experiment 2) when observers were not told of the to-be-reported demographic property (either perceived race or gender) until after the stimulus has disappeared — thus suggesting that there is no cost to extracting two different demographic variables compared to only one. Second, we observed (in Experiment 3) that the temporal profile of perceiving demographic properties in faces follows a similar pattern to the detection of faces in the first place. In both cases, performance was at chance at 16.66 ms, was above chance by 33.33 ms, and was impressively good by 50 ms. (And in contrast, performance without the masks was dramatically improved for both the demography and detection tasks, thus demonstrating the masks' effectiveness in the first place.)

Note that these experiments were designed in order to test the *minimal* exposure time that is required to extract demographic properties, which may differ (and be shorter) than the degree of exposure that we *typically* employ during free viewing. In particular, this goal entailed a series of design choices that are incompatible in some ways with explorations of how fast we *typically* extract demographic properties from faces. For example, exposure time was intentionally blocked in all experiments reported here: this is especially useful for testing questions about *minimal* exposure times, but of course we do not typically encounter faces that are blocked in this way during everyday life. (Intermixed exposure durations would be more akin to everyday experience, but would not allow us to ask how fast we *can* extract demography from faces.) Similarly, the wide variety of demographic traits that we encounter in everyday life was simplified in the current experiments in terms of both the stimuli (which were chosen to be maximally discriminable along dimensions of interest) and the response options (which were 2AFC tasks: White/Black, Young/Old, and Female/Male). As such, these results are informative with respect to the *minimal* exposure time required to extract demography, but it remains unclear whether or how they might generalize to more ambiguous stimuli (e.g. middle-aged faces) or more complex decisions (e.g. when estimating an exact age rather a categorical age). And similar points apply to other design choices, such as standardizing the presentation locations.

The efficiency of perceiving demography from faces has several theoretical implications. First, and perhaps most importantly, these results suggest that demographic features are a foundational part of face perception itself — and at a minimum that they are extracted no less efficiently than are other properties (such as extraversion and aggressiveness) that have received considerably more attention in research on speed and efficiency. Second, we have frequently referred to “demographic properties” as a unitary class in our discussion, and it was especially noteworthy that both perceived race and perceived gender produced the same patterns of results. This finding is in stark contrast to predictions that perceived race and perceived gender should diverge dramatically, due to the fact that the latter — but not the former — was especially salient in our evolutionary past (Cosmides et al., 2003; Kurzban et al., 2001; Pietraszewski et al., 2014). Finally, these results have important implications for the use of temporal constraints in face perception research in general. In some such studies, researchers have attempted to explore what kinds of effects (such as the influence of perceived race and emotion on recognition accuracy) will occur in a “constrained condition” where not much exposure is available — but that exposure may be as long as, say, 500 ms at its very *shortest* (e.g. Ackerman et al., 2006). And while 500 ms may indeed sound like an amazingly brief duration when considering the temporal scale of everyday life, the current results illustrate that the perception of demography requires less than a fifth of that exposure to be as effective as can be.

Author Note

For helpful conversations, we thank Yi-Chia Chen, Mike Morais, Alexander Todorov, and the members of the Social Perception laboratory and the Perception & Cognition laboratory.

Declarations

Funding: This project was funded by ONR MURI #N00014-16-1-2007 awarded to BJS.

Conflicts of interest/Competing interests: The authors have no relevant financial or non-financial interests to disclose.

Ethics approval: This study was performed in line with the principles of the Declaration of Helsinki. All experimental methods and procedures were approved by the Yale University Institutional Review Board.

Consent: Informed consent was obtained from all individual participants included in the study.

Availability of data, materials, and code: Preregistrations, raw data, and code for each experiment can be viewed at https://osf.io/pzrhw/?view_only=c85385291bf04d278e5e6609034b17ec

Authors' contributions: C. Colombatto, S. Uddenberg, and B. J. Scholl designed the experiments; C. Colombatto conducted the experiments; C. Colombatto and S. Uddenberg analyzed the data; C. Colombatto, S. Uddenberg, and B. J. Scholl wrote the manuscript.

References

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., Maner, J. K., & Schaller, M. (2006). They all look the same to me (unless they're angry). *Psychological Science*, *17,* 836-840.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*, 269-278.

Barragan-Jason, G., Lachat, F., & Barbeau, E. J. (2012). How fast is famous face recognition? *Frontiers in Psychology*, *3*:454, 1-11.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551-565.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, *100*, 3-18.

Besson, G., Barragan-Jason, G., Thorpe, S. J., Fabre-Thorpe, M., Puma, S., Ceccaldi, M., & Barbeau, E. J. (2017). From face processing to face recognition: Comparing three different processing levels. *Cognition*, *158*, 33-43.

Besson, G., Ceccaldi, M., Didic, M., & Barbeau, E. J. (2012). The speed of visual recognition memory. *Visual Cognition*, *20*, 1131-1152.

Borkenau, P., Brecke, S., Mötting, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality, 43,* 703-706.

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Attitudes and Social Cognition, 88,* 885-894.

Colombatto, C., & McCarthy, G. (2017). The effects of face inversion and face race on the P100 ERP. *Journal of Cognitive Neuroscience, 29,* 664-676.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, *37*, 1102-1117.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7,* 173-179.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades towards faces: Face detection in just 100 ms. *Journal of Vision, 10*:4, 1-17.

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications, 10*:1258, 1-10.

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science, 27,* 502-517.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, *14*, 251-262.

George, P. A., & Hole, G. J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition*, *7,* 485-509.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1023-1039.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152-160.

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113,* 513-529.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98,* 15387-15392.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122-1135.

Mark, L. S., Pittenger, J. B., Hines, H., Carello, C., Shaw, R. E., & Todd, J. T. (1980). Wrinkling and head shape as coordinated sources of age-level information. *Perception & Psychophysics, 27*, 117-124.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *1,* 3-35.

Minear, M., & Park, D.C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*, 630-633.

O'Toole, A. J., Vetter, T., Volz, H., & Salter, E. M. (1997). Three-dimensional caricatures of human heads: Distinctiveness and the perception of facial age. *Perception, 26*, 719-732.

Peirce, J. W., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195-203.

Pietraszewski, D., Cosmides, L., & Tooby. J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLOS One*, *9*:2, e88534.

Pietraszewski, D., Curry, O. S., Petersen, M. B., Cosmides, L., & Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition*, *140,* 24-39.

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2011). Holistic processing of faces happens at a glance. *Vision Research, 49,* 2856-2861.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research, 51,* 333-342.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews, 1,* 63-75.

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*, 440-455.

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

Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005). Towards realism in facial prototyping: results of a wavelet MRF method. *Theory and Practice of Computer Graphics*, *24*, 449-456.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton, NJ: Princeton University Press.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519-545.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471-491.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42,* 671-684.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592-598.

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology, 117,* 364-385.

Xie, S. Y., Flake, J. K., Stolier, R. M., Freeman, J. B., & Hehman, E. (in press). Facial impressions are predicted by the structure of group stereotypes. *Psychological Science.*

Figure Captions

Figure 1. Examples of (a) the face stimuli and (b) the masks used in these experiments. (Note that due to restrictions on reproductions of stimuli from the face databases, the stimuli for the age task — from the left column of panel a — are similar to but not identical to the ones employed in the experiment.)

Figure 2. An overview of the experimental procedures in Experiments 1-3. (a) In Experiment 1, observers viewed a face either for a limited duration (ranging from 16.66 to 1000 ms; top row), or for as long as they wished (bottom row). (b) In Experiment 2, observers were informed of the demographic property to be queried either before the face was displayed (top row), or after the face was no longer visible (bottom row). (c) In Experiment 3, observers viewed either an unscrambled face (Face Present trials) or a different scrambled face (Face Absent trials), followed by either a mask (Masked trials) or a blank screen (Unmasked trials).

Figure 3. Results from Experiment 1: Average correlations (presented along the vertical axis) between responses (of perceived race, age, and gender) to the same stimulus viewed at several constrained exposure durations (presented along the horizontal axis) and either (a) responses with unlimited exposure, or (b) the ground truth. Error bars reflect 95% confidence intervals on the raw correlations after subtracting out the shared variance. (As noted in the main text, these results were unaltered using the Fisher-Z transformed values.)

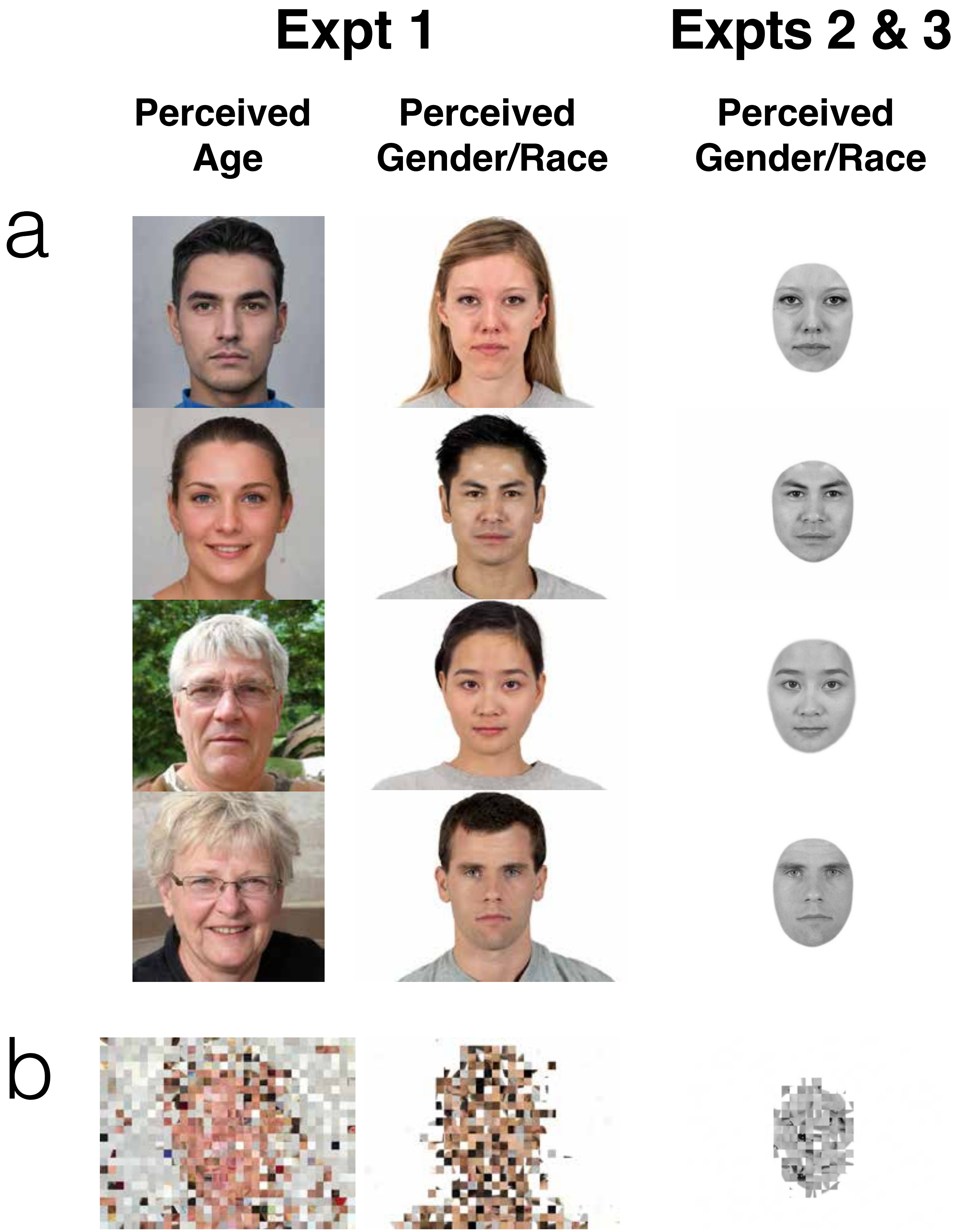
Figure 4. Percentages of raw responses (on a scale from 1 to 5) for the various perceived demographic properties at several exposure durations in Experiment 1. Observers used the extreme keys ('1' or '5') to indicate that the face was seen as "definitely" White (or Black, or Young, or Old, etc.), and they used the middle keys (e.g. '3') when they were uncertain. These

results clearly indicate that observers used only the extreme ("definitely") responses at exposures of 50 ms or above, and only used the uncertain responses at 16.66 ms.

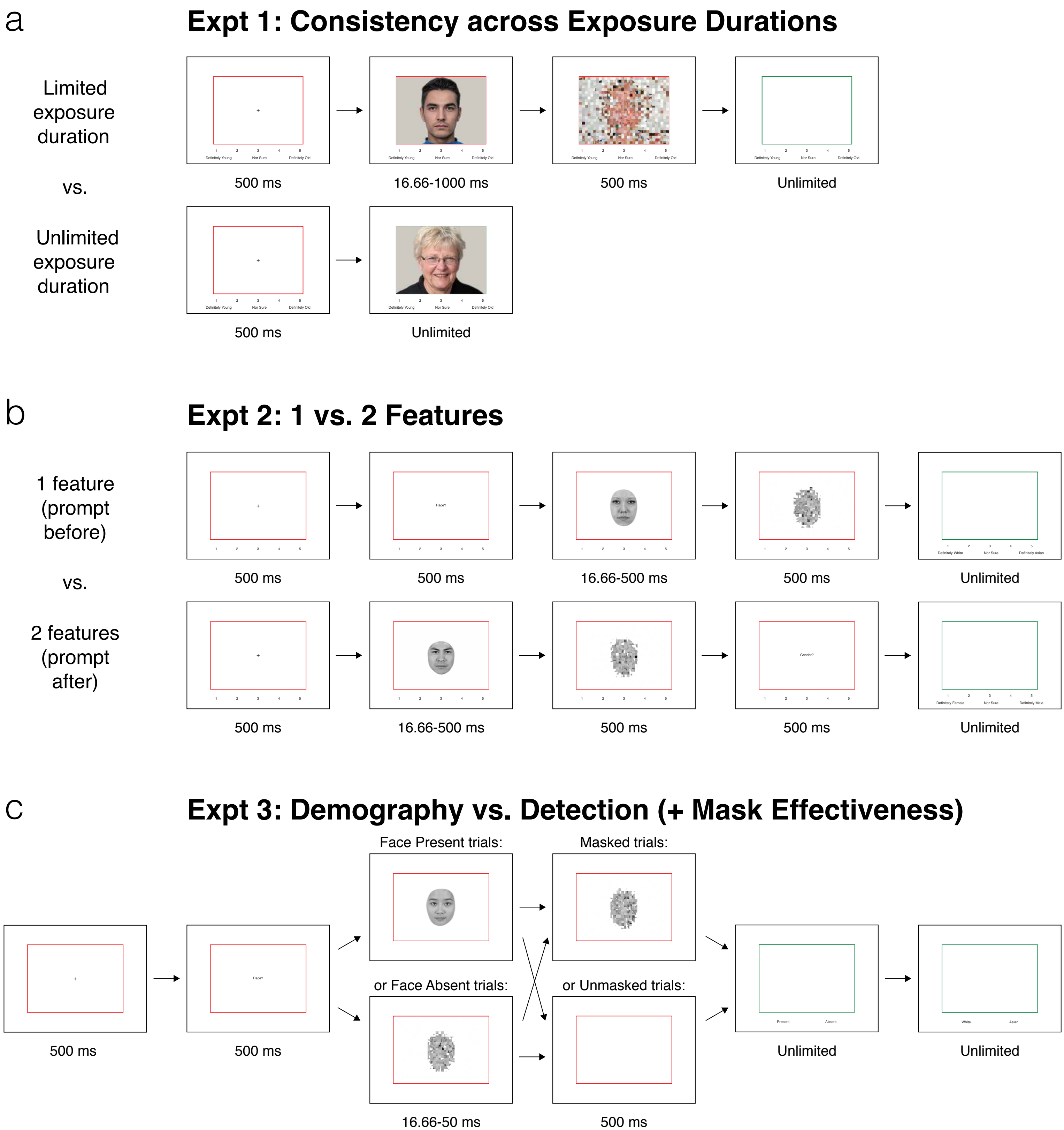
Figure 5. Results from Experiment 2. Average correlations (presented along the vertical axis) between responses (of perceived race or gender) to the same stimulus viewed at several constrained exposure durations (presented along the horizontal axis) and the same responses with unlimited exposure. Solid lines reflect blocks wherein the to-be-reported property was identified in advance; faded lines reflect blocks wherein the to-be-reported property was not revealed until after the face had disappeared. Error bars reflect 95% confidence intervals after subtracting out the shared variance.

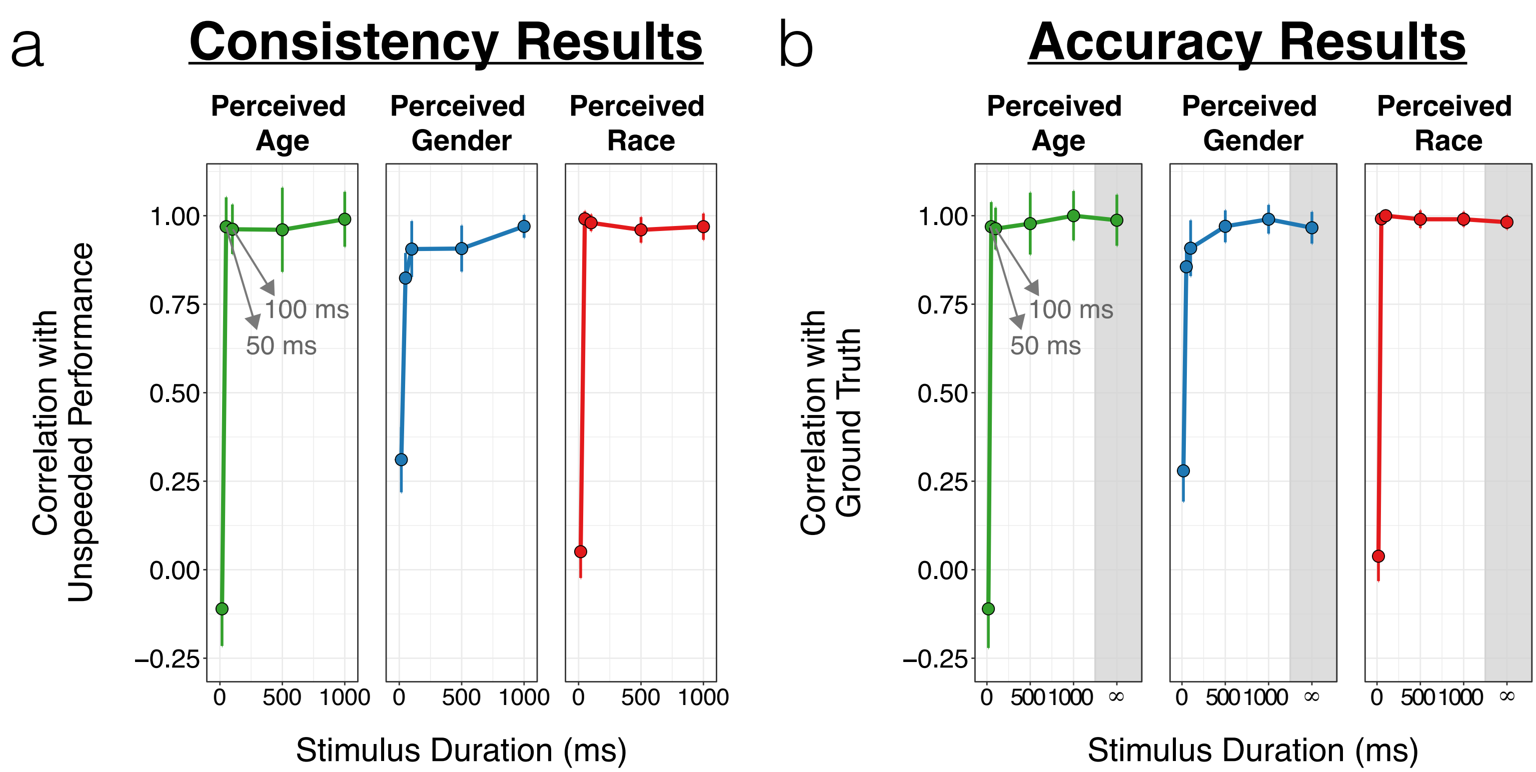
Figure 6. Results from Experiment 3. (a) Average correlations (presented along the vertical axis) between responses (of perceived race or gender) to the same stimulus viewed at several constrained exposure durations (presented along the horizontal axis) and the same responses with unlimited exposure. (b) Average face detection sensitivity (in terms of d', presented along the vertical axis) for several constrained exposure durations (presented along the horizontal axis). Solid lines reflect blocks with masked faces; dashed lines reflect blocks without any masks. Error bars reflect 95% confidence intervals after subtracting out the shared variance.

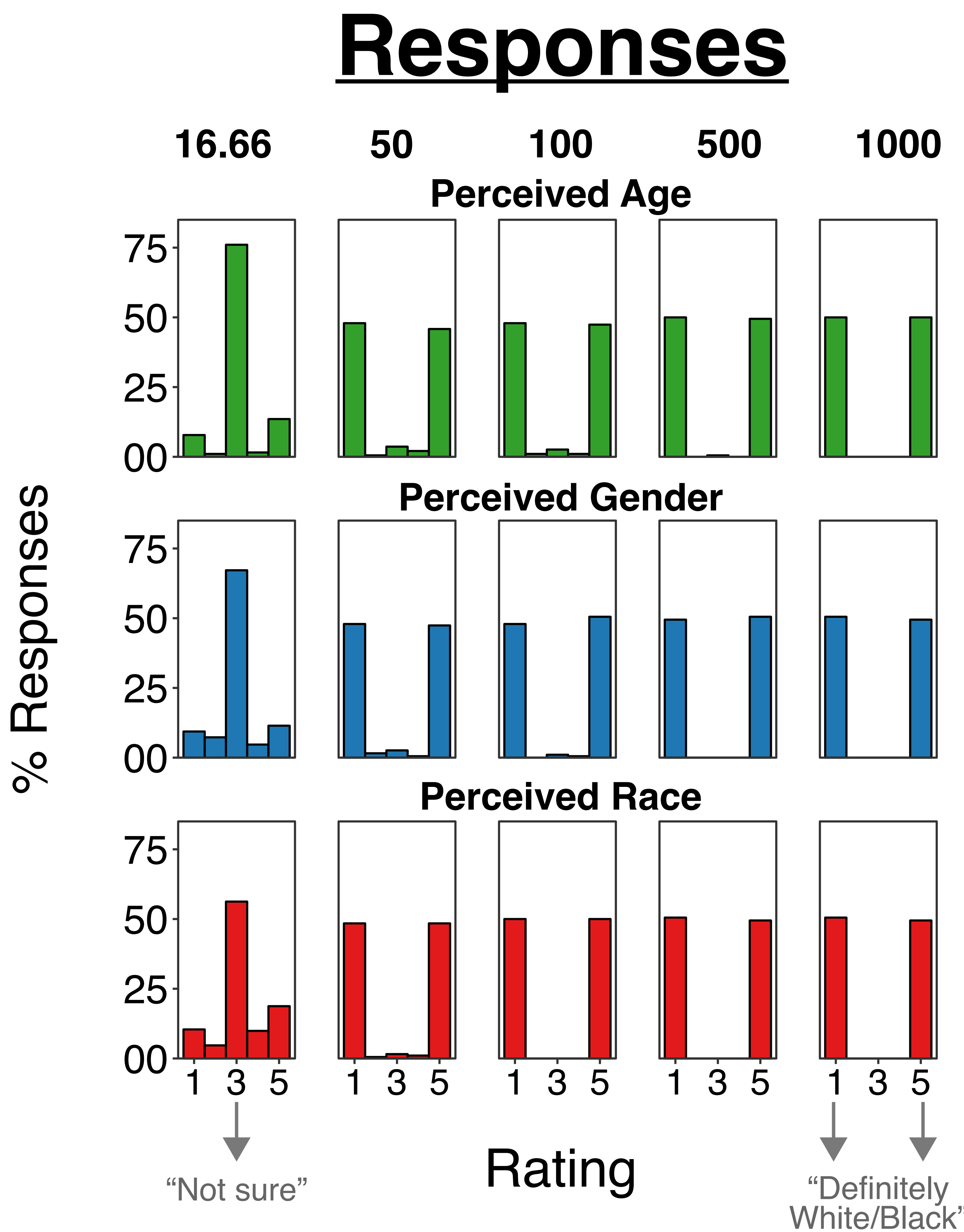
Sample Stimuli



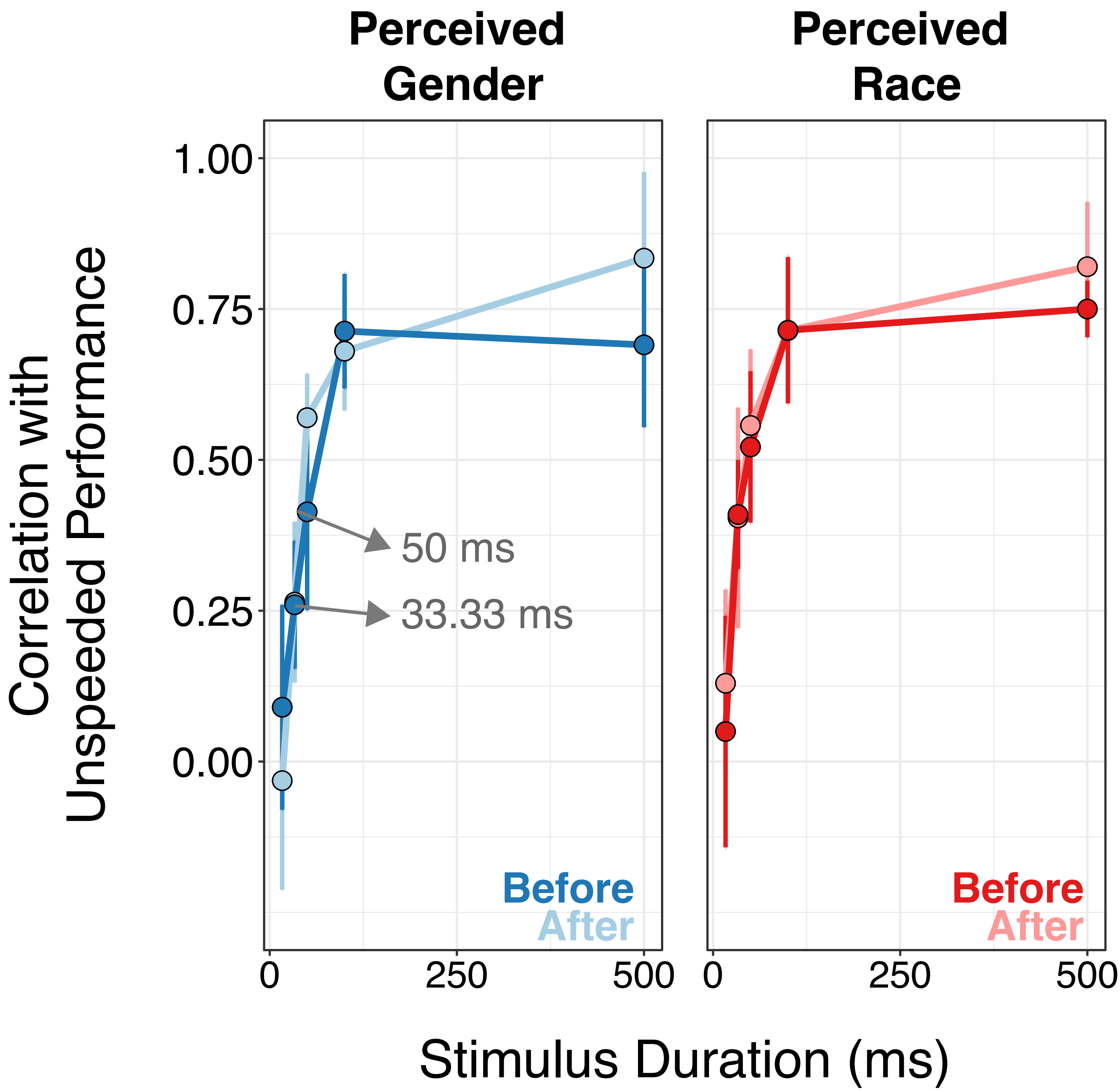
Experimental Procedures

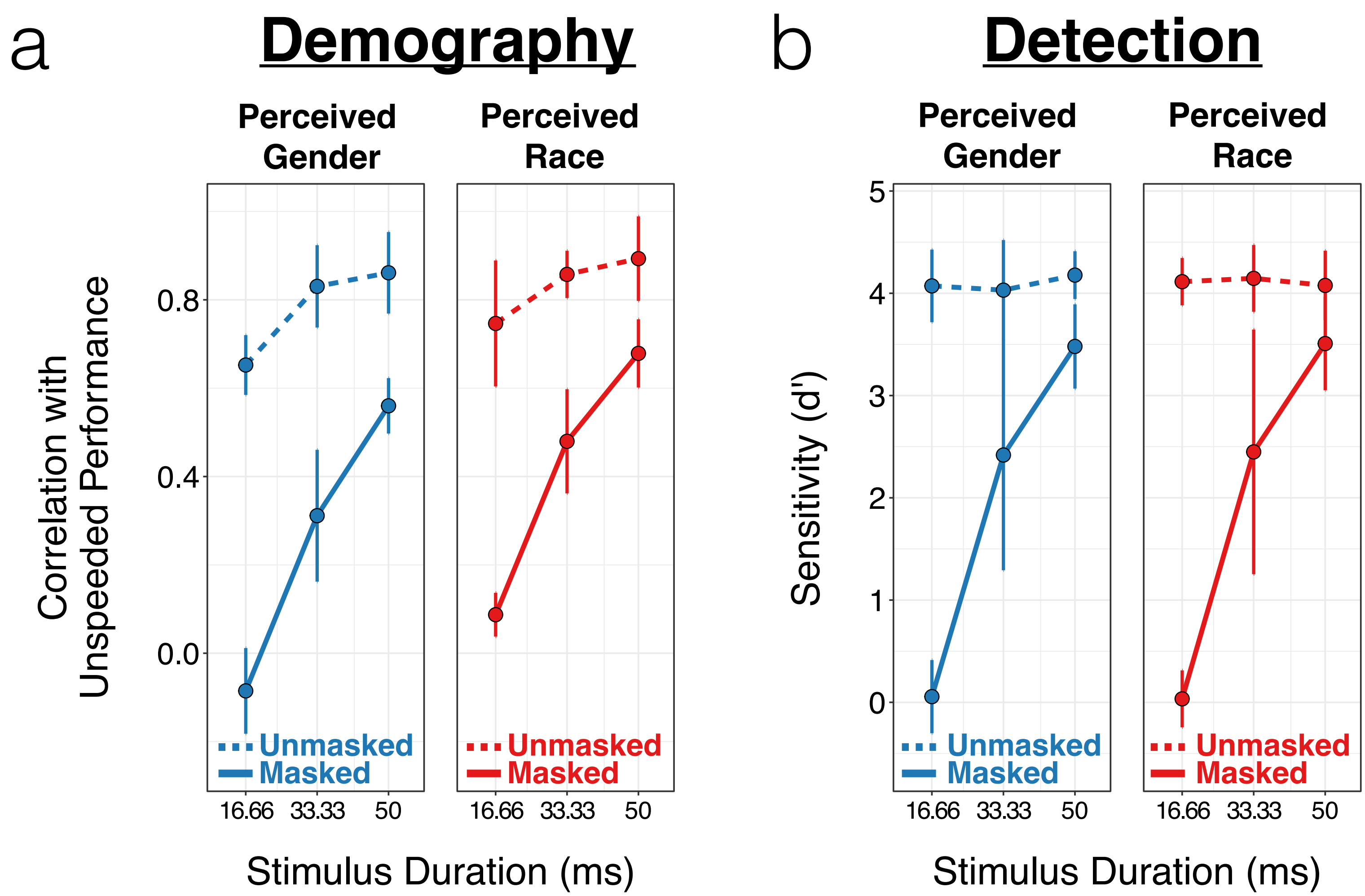



a
Consistency Results
Perceived Age
Perceived Gender
Perceived Race
Correlation with Unspeeded Performance
100 ms
50 ms
1.00
0.75
0.50
0.25
0.00
−0.25
0
500
1000
Stimulus Duration (ms)
b
Accuracy Results
Perceived Age
Perceived Gender
Perceived Race
Correlation with Ground Truth
100 ms
50 ms
∞
Stimulus Duration (ms)


Responses
16.66
50
100
500
1000
Perceived Age
Perceived Gender
Perceived Race
% Responses
75
50
25
00
1
3
5
Rating
"Not sure"
"Definitely White/Black"

1 vs. 2 Features




a
Demography
Perceived Gender
Perceived Race
Correlation with Unspeeded Performance
0.0
0.4
0.8
Unmasked
Masked
16.66
33.33
50
Stimulus Duration (ms)
b
Detection
Perceived Gender
Perceived Race
Sensitivity (d')
0
1
2
3
4
5
Unmasked
Masked
16.66
33.33
50
Stimulus Duration (ms)