NORTHWESTERN UNIVERSITY

Where Do We Come From? What Are We? Where Are We Going?

Contemplating Artificial Intelligence Applications in Organizations and Organizational Research

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

in Management and Organizations

By

Dawei Wang

EVANSTON, ILLINOIS

September 2022

Abstract

"Where Do We Come From? What Are We? Where Are We Going?" is the name for one of French artist Paul Gauguin's most influential paintings. Unsurprisingly, these very questions have occupied the minds of countless philosophers, artists, and scholars since the beginning of human civilization. These questions become especially salient when drastic changes occur in our environment, such as pandemics, wars, global economic challenges, and disruptive technological advancements. In a rare coincidence, humankind is faced with all these challenges at this point in time. Thus, this dissertation humbly contemplates these important questions, not only in the context of organizations and the future of work, but life in general.

As artificial intelligence is applied increasingly in our lives, changing the way we live, work, and play, organizations and organizational research arrive at a juncture where their participants and members must ask: "Where do we come from? What are we? Where are we going?" If artificial intelligence is becoming as omniscient as the rational demons described in most economic research, where does it leave management and organizations as a field as well as existing organizations in the field? Perhaps a more fitting and realistic question is: "Is artificial intelligence as powerful as we imagined it to be?" Or, should we adopt a normative lens and paint a blueprint for future researchers, policy-makers, and other people in the world? Should we help navigate the relationship with machines in the inevitable applications of artificial intelligence in people's lives?

The first two chapters of this dissertation are empirical. They go deep into the technical aspects of existing artificial intelligence algorithms, and explore the limits and capabilities of artificial intelligence technology. I found that artificial intelligence tools, at the moment, are not as powerful as we imagined. Shortcut learning and biases, as well as misinterpretation of data and results, are just some of the issues I observed through my research. The last chapter attempts to answer some normative and theoretical questions. It draws upon works by pioneering researchers in both artificial and organizational

intelligence research, and provides a working lens or framework for how we can make sense of the currently fragmented and noisy landscape in artificial intelligence application research.

As scientific as it strives to be, this dissertation, in my opinion, should be more fittingly viewed as a faith declaration and an expression of my belief that organizations and the human members therein are bigger than the current artificial intelligence phenomenon. In declaring my faith, I hope my research contributes to a more human-centered direction of where we can go as a field in the wake of artificial intelligence technology. The dissertation will also help the organizational field will also consider the organizational element of artificial intelligence application, that is, how this technological development can be integrated as a larger organizational phenomenon.

Acknowledgments

It has been a great honor and privilege to undergo my doctoral training at Northwestern University and the Northwestern Kellogg School of Management, during which I was also a part of the Northwestern Institute on Complex Systems and the Amaral Lab, and took valuable classes in the Northwestern Pritzker School of Law and the Northwestern McCormick School of Engineering. My research interest is interdisciplinary, as is Northwestern and my training, but I would not be successful in crafting this unique identity without the help of my advisor, dissertation committee members, and professors: Edward Zajac, Luis Amaral (and colleagues at the Amaral Lab), Maryam Kouchaki, Hatim Rahman, Brian Uzzi (and colleagues at the Northwestern Institute on Complex Systems), Adam Pah, Dashun Wang, Hyejin Youn, and Zhaoli Song (and colleagues at the National University of Singapore).

I would also like to give special thanks to my collaborators for pursuing interesting research projects with me, Xiuxi Zhao, Anyi Ma, Qinghua Li, Xiaoyun Xie, Fan Zhou, Krishnan Nair, and colleagues from Zhejiang University, as well as to people and groups who gave me technical support, Ron Nowak, Amirsina Torfi, Rohit Deo, Northwestern Quest, and Kellogg Research Support. Additionally, I would like to thank the editors and reviewers of my several papers, Krishna Savani, Shinobu Kitayam, and Michal Kosinski, who taught me valuable lessons in the publishing process.

The doctoral program was made meaningful by my cohort mates, Hannah Birnbaum, Anna McKean, Hui Sun, Hannah Waldfogel, and Amber Johnson, as well as by my doctoral program mates, Haochi Zhang, Ming Wang, Wei Wang, Binglu Wang, Yuan Tian, and Yian Yin. I managed to stay relatively healthy because of my basketball friends, Derek Nkemnji, Tre Wells, Jon Pilarski, and Tem Gebrekristos, and the caring doctors at Northwestern University Healthcare, who treated my hypertension. I would like to thank my wife, Zhuoli Shen, whom I met because of Northwestern Pritzker School of Law and who truly believed in me, and my daughter, Loralei Grace Wang, who was born during the fourth year of my doctoral program and gave me inspirations for my research. I deeply appreciate my family and

friends who stayed with me during the highs and lows of my doctoral program, my wife's grandmother, Yuanfang Chen, my pastor, Daoyuan Tang, our beloved neighbors, Jennifer Drake and Eric Anderson, my best friends, Sean Yuan, Jerry Chai, and Yonggang An, and my parents, Gang Wang and Mei Xiao.

This dissertation is dedicated to my advisor, Edward Zajac, who inspired and supported me.

Table of Contents

List of Tables and Figures

Introduction

Artificial intelligence technology is making tremendous breakthroughs. According to Moore's law, the number of components in semiconductors has been growing exponentially each year, and will continue to do so in the years to come[1] (Schaller, 1997). With more components, computers can process information much faster and cheaper than a few decades ago. Computers can also store data more efficiently, reliably, and economically. Simultaneously, sophisticated software has been invented to capitalize on these hardware developments, such as deep neural network algorithms that mimic the structure of the human brain (Netzer et al., 2011). These algorithms consolidate, find patterns, and generate valuable insights from unstructured data, allowing humans to make sense of information, leading to better decisions (Choudhury et al., 2021).

Thanks to these developments, artificial intelligence has accomplished some extraordinary feats. For example, artificial intelligence defeated the best-ranked humans in games such as Go (Silver et al., 2017) and Jeopardy (Ferrucci et al., 2013). Self-driving cars, which rely on artificial intelligence, are becoming more autonomous and commonplace (Badue et al., 2021). Tiku (2022) even suggested that artificial intelligence has finally reached sentient level, based on a Google engineer's claim that Google's chatbot came to life and was able to comprehend emotions and understand its rights, on top of conversing fluently with humans.

With these accomplishments, more organizations are starting to adopt artificial intelligence in their services and internal management systems. For example, in the banking industry, organizations are using chatbots in their customer service, significantly replacing human customer service officers

[1] Although experts have observed that the speed of growth has decreased in recent years, the efficiency and energy-saving abilities keep improving.

(Adamopoulou & Moussiades, 2020). Some companies, such as Charles Schwab,[2] are also employing voice and facial recognition to verify the identity of their customers (Aravinda et al., 2022). In negotiation contexts, researchers believe that the use of artificial intelligence can help customers negotiate better deals (Dai et al., 2021). Some start-ups, such as Intellext,[3] are capitalizing on this idea, providing negotiation platforms where artificial intelligence negotiate deals on behalf of humans. In employee facing functions, companies, such as Enaible[4], provide productivity monitoring systems that track employee progress, consolidate information, and intelligently provide positive habit-building recommendations and well-being suggestions for employees (Pan & Zhang, 2021). Human resource vendors, such as HireVue[5], utilize artificial intelligence algorithms to quickly interview and screen job candidates, saving time and resources for human resource managers (Peña et al., 2020).

Turning to academia, there is much discourse about how artificial intelligence can be applied to research. Table 1 presents a non-exhaustive list of examples in behavioral research that applied artificial intelligence tools. In theory building, Leavitt et al. (2020) argued that artificial intelligence can be applied to research because it can test mid-level theories that are otherwise infeasible to test using traditional methods. Csaszar and Steinberger (2022) proposed that organizational theorists can borrow ideas from artificial intelligence research because organizations and artificial intelligence are very similar. Social psychological researchers argue that artificial intelligence algorithms can assist researchers in generating novel hypotheses (Sheetal et al., 2020; Sheetal & Savani, 2021). Researchers can employ algorithms in preliminary studies, find interesting and novel patterns in big data, and replicate these results using lab studies (Sheetal et al., 2020; Sheetal & Savani, 2021).

---

[2] https://www.schwab.com/
[3] https://www.intellext.ai/
[4] https://enaible.io/
[5] https://www.hirevue.com/

----- INSERT TABLE 1 ABOUT HERE -----

Apart from theory building, a number of research papers have been published in recent years utilizing artificial intelligence as a key component of their methodology. In the past, facial perception researchers relied on human research assistants to extract facial features (e.g., Ambady & Rosenthal, 1993; Rule & Ambady, 2010). Researchers asked assistants to manually label variables in their database, such as whether or not the participant in each facial image is smiling. With artificial intelligence, Wang et al. (2019) passed their facial image data through an application programming interface (API), an interface that allows the researchers to interact with the artificial intelligence algorithm, and extracted facial variables, such as smiling, for their research. Wang et al. (2019) found that artificial intelligence-based labels predicted similar results compared to human raters. In another study, instead of experimenting with humans in the laboratory as in traditional research, Wang (2021) conducted image distortions and experimented directly with the facial recognition algorithm to find possible mechanisms driving the relationship in his study.

While there are many opportunities to capitalize on the developments and accomplishments of artificial intelligence, application of the technology in organizations and research is met with many challenges. More and more evidence is surfacing, demonstrating how artificial intelligence would become biased or easily fail. One report, for example, showed that Tesla's self-driving systems are, in fact, causes of car crashes (Boudette et al., 2022). The aforementioned artificial intelligence-based hiring system by HireVue received a federal complaint and was later halted because the system was believed to be biased against minorities (Harwell, 2019). Similarly, Amazon halted its resume screening algorithm because it was biased against women (Dastin, 2018). The list of examples where artificial intelligence is either biased or harmful goes on.

Indeed, the problem of artificial intelligence algorithms discriminating against minorities is now widely established in research (Drozdowski et al., 2020). Scholars interested in algorithmic bias, which is the study of how algorithms would systematically discriminate against certain sub-groups, have found that artificial intelligence would easily pick up biases in training data, perpetuate the bias in its parameter optimizing process, and behave in biased ways when implemented in the field (Suresh & Guttag, 2021). For example, in facial recognition, because there are less photographs in the training data for younger people, the prediction accuracy for younger people is much lower than for adults (Drozdowski et al., 2020).

Apart from algorithmic biases, researchers have found that algorithms are actually not capable of learning objective features in their respective tasks (Geirhos et al., 2020). In this phenomenon, called shortcut learning, machines rely on shallow or superficial features to make predictions (Geirhos et al., 2020). For example, in image recognition, an algorithm would mistakenly label a photograph of greenery and clouds—"sheep." This is because sheep and greenery often occur together in images, and thus when the algorithm was "optimizing" its parameters for sheep, it would "lazily" associate features representing greenery with features for sheep, instead of using the shape or texture of the sheep to predict sheep in images (Geirhos et al., 2020). This problem does not only exist in image recognition, but virtually all domains of artificial intelligence.

Given these opportunities and challenges in artificial intelligence applications, scholars need to answer two pressing questions: "should and how should artificial intelligence be managed in organizations and organizational research?" and "how should we shape our path forward?" The first question is targeted at balancing the optimistic and negative view of artificial intelligence application. Advocates of artificial intelligence often imagine a utopian society enabled by artificial intelligence, while critics propose a dystopian society. My opinion, as shown in this dissertation, is that this question is both descriptive and prescriptive. A deeper understanding of artificial intelligence would provide us a realistic

description of the capabilities and limitations of artificial intelligence. Based on this understanding, scholars should then prescribe appropriate research, policy, or managerial recommendations to guide the management of artificial intelligence application.

In this dissertation, I address these questions by looking at both the methodological and theoretical aspects of artificial intelligence application. Given the technical nature of artificial intelligence, an understanding of these questions must be motivated by a deep comprehension of the inner-workings of these systems. For example, to study the application of facial recognition systems in social psychological research, researchers must at least have a deep grasp of the entire system—how it is designed, trained, and implemented and how its results are interpreted. Going deep, I conducted a series of experiments, which meticulously dissected each step in the pipeline of artificial intelligence application. These experiments are reported in the first two chapters.

Armed with such understanding, I went broad and theoretical, asking questions such as "how should we conceptualize artificial intelligence?" and "how should we integrate artificial intelligence in future organizational research?" These theoretical questions are discussed in the third chapter, which paints a theoretical framework of how organizations and organizational researchers could more appropriately manage the application of artificial intelligence. The chapter advocates a more human-centered and organizational perspective.

By going deep and wide on the topic of artificial intelligence application, I hope this dissertation can help us understand where we come from, what we are, and where we are going as scholars of the phenomenon of artificial intelligence application. Perhaps, with this better understanding, we can build a better future for organizations, organizational research, and what may come beyond.

Chapter 1

Presentation in Self-Posted Facial Images Can Expose Sexual Orientation: Implications for Research and Privacy

Introduction

Several recent studies have found that sensitive personal attributes are becoming increasingly easy to detect using facial images. Advanced facial recognition algorithms can now accurately classify sensitive traits, such as sexual orientation (Wang & Kosinski, 2018), personality (Kachur et al., 2020; Wolffhechel et al., 2014), political orientation (Kosinski, 2021), and unlawful behaviors (Wu & Zhang, 2016. For example, Wang and Kosinski (2018) found that an off-the-shelf facial recognition algorithm can be easily repurposed into a sexual orientation classifier that can differentiate sexual orientation with a classification rate[6] of above 80% for men and 70% for women from a single naturalistic facial image, considerably more accurate than what can be achieved by human judges. In another study, a similar off-the-shelf algorithm, using Facebook and dating profile images, was shown to classify individuals' political orientation with a classification rate of over 70% (Kosinski, 2021).

What is unclear is to what extent the classifications were driven by fixed (i.e., facial morphology), transient (i.e., grooming styles), and non-facial (e.g., background or lighting) image features. Facial recognition research posits that if faces were aligned at the same position in the facial images used to train the algorithm, each pixel in the image would map onto a specific facial feature (Parkhi et al., 2015; Taigman et al., 2015). Similarly, Taigman et al. (2015) referred to their algorithm as

[6] Classification rate is expressed as area under the receiver operating characteristic curve (AUC). When presented with stimulus X from category A and stimulus Y from category B, the AUC refers to the extent to which the model assigns Y a higher probability of belonging to category B than X.

"a well localized description of the underlying face" because they found that pixels that activated the algorithm were in the facial area. Building on these findings, Wu and Zhang (2016, p. 2) claimed that "sophisticated algorithms based on machine learning may discover very delicate and elusive nuances in facial characteristics and structures that correlate to innate personal traits." Stoker et al. (2016, p. 8) even described machine learning as an "advanced objective method for the measurement of facial features."

On the other hand, some researchers acknowledge the contribution of both facial morphology and grooming features. For example, Wang and Kosinski (2018) explained that "[f]acial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style)" (p. 246), though they traced some of these differences to biological predispositions, citing the prenatal hormonal hypothesis (e.g., "According to the PHT [prenatal hormonal theory], same-gender sexual orientation stems from the underexposure of male fetuses or overexposure of female fetuses to androgens that are responsible for sexual differentiation" (p. 247)). Kachur et al. (2020) claimed that "machine learning… could reveal multidimensional personality profiles based on static morphological facial features" but the researchers were "still unable to claim that morphological features of the face explain all the personality-related image variance captured by the ANNs" (p. 6). Agüera y Arcas et al. (2018) surveyed 8,000 Americans of different sexual orientations and asked them to fill out an array of "yes/no" questions about their self-presentational style. The results showed that gay subjects were more likely to report wearing glasses and less likely to report having face tan and working outdoors.

These findings are consistent with the impression management literature, which posits that people intentionally or unintentionally shape how they are seen by others (Goffman, 1959; Leary & Allen, 2011; Leary & Kowalski, 1990; Schau & Gilly, 2003; Schlenker, 2012). For example, people choose to present themselves differently by constructing distinctive text-based self-descriptions on online

dating platforms (Schau & Gilly, 2003; Tong et al., 2020). Self-presentational studies also found that in unfamiliar and different-sex social interactions, people felt more nervous, thought about others' impressions more, and wanted to make better impressions than they did in familiar and/or same-sex interactions (Leary et al., 1994). When sexual orientation comes into play, gay and heterosexual men have distinctive aesthetic appeals (Rudd, 1996). Gay men on average preferred innovative or trendy apparel, whereas heterosexual men preferred casual or laid-back styles.

The idea that people choose different photographs depending on the context is not new in facial perception literature. Todorov and Porter (2014) explained that "[w]ebsite users did not randomly select which images of themselves to post on these sites. Hence, it is possible that the presumed accuracy reflects biases in the selection of the images rather than honest or inherent signals of sexual orientation in the face" (p. 1415). The authors asked participants to select images of different facial expressions and found that in scenarios such as dating, participants chose photographs that portrayed a trustworthy-looking face compared to a mean-looking face of the same person. These findings were confirmed by White et al. (2017), who asked participants to upload 12 photographs of themselves to a professional and a dating website. Participants systematically uploaded more attractive-looking photographs to the dating website and more competent-looking photographs to the professional website. Hancock and Toma (2009) found that in online dating photographs, people engaged in self-enhancement tactics to make themselves appear more attractive. These tactics varied by gender; women posted photographs of themselves taken when they were younger, and were more likely to re-touch their photographs compared to men.

Understanding how presentation in self-posted facial images influences classification of sexual orientation is critical to the ongoing discussion about privacy (Matz et al., 2020). If trait classifications were mostly driven by morphological differences, as implied in existing facial recognition research, privacy loss would be preventable using existing de-identification technologies. For example, consumers

can protect themselves using do-it-yourself data protection software that masks, blurs, or pixelates the facial regions of these images, or by wearing face masks (Li et al., 2017; Matzner et al., 2016; Shan et al., 2020; Zhang et al., 2014). However, if classifications were also driven by self-presentation, the danger of privacy loss might be greater than previously believed. There are more dimensions on which self-presentation may vary. Pinpointing the exact features in the facial image where private information is retrieved would be difficult. In this case, the burden of privacy protection must be shifted to governments and companies because the alternative to privacy protection would be to ask consumers to refrain from self-expressions (i.e., not post images online), a tradeoff that many might not want to make.

To understand how self-presentation influences the ability to extract sexual orientation information from self-posted photographs, I collected a dataset consisting of 15,286 gay and heterosexual participants from an online dating website. I first obtained 12 self-presentational facial attributes from their facial images, tested whether there were significant differences according to sexual orientation (Study 1a), and examined the degree these differences contributed to classification of sexual orientation (Study 1b). Then, I replicated the sexual orientation classification algorithm. I tested the contribution of image background on the classification of sexual orientation (Study 1c). I masked the facial portion in each facial image so that only background information remained. If masked images, in which only the image background was retained, can classify sexual orientation, it means that the image background (a self-presentational feature) contributed to sexual orientation classification.

Next, previous research suggests that gay men appear to have brighter skin tone compared to heterosexual men. I tested whether skin tone was related to the overall brightness of the face and/or the background (Study 2a). If skin tone was related to overall brightness (both face and background), it is likely that gay vs. heterosexual people presented themselves in images with varying levels of brightness or illumination. Finally, I tested the contribution of overall image brightness on sexual

orientation classification (Study 2b). I blurred each facial image so that only three numbers representing the brightness of each color channel remained. If a completely blurred out image can classify sexual orientation, then people's choices of image brightness, or the illumination of their ambient environments (indoors vs. outdoors), contributed to sexual orientation classification.

Study 1a

Agüera y Arcas et al. (2018) found that reported self-presentational styles varied by sexual orientation. Some self-presentational differences were also reported by Wang and Kosinski (2018), such as the likelihood of wearing glasses in facial images. Study 1a aimed to extend their findings and test the extent these differences are observed in self-posted facial images on a dating website.

Methods

Preprocessing Facial Images. All data collection was conducted after the study was approved by the institutional review board (IRB) of my university. Following the exact procedures described in Wang and Kosinski (2018), I collected facial images of public profiles from a U.S. dating website. I aimed to collect a larger sample than the previous study to achieve higher generalizability. I gathered a total of 76,181 profiles (412,446 images), of which 39,386 were women (224,855 facial images) and 36,795 were men (187,591 facial images), aged 18 to 40.

Next, I cleaned and preprocessed all facial images with the help of the Face++ API, which is a facial recognition software widely used in facial research (Kosinski, 2021; Wang et al., 2019; Wang & Kosinski, 2018), verified to be accurate at extracting facial information from images (Jaeger et al., 2020). Four sets of information were extracted: the number of faces in each facial image, facial landmarks, facial attributes, and facial bounding boxes. Like Wang and Kosinski (2018), I dropped images that did not contain human faces, contained more than one face, contained partially hidden faces (i.e., if any occlusions were detected by the Face++ API), or had small-resolution faces (i.e., the width of the bounding box was less than 40 pixels). Like the previous study, I also removed images in which faces

were significantly turned away from the camera (i.e., head pose yaw angle greater than 15 degrees and pitch angle greater than 10 degrees). All these steps were taken to ensure that every face in the dataset would align in the same position and angle, allowing the facial recognition algorithm to accurately recognize the face.

Following Wang and Kosinski (2018), I only included Caucasian individuals aged 18 to 40. However, instead of verifying the demographics manually, I relied on the gender and age detector of Face++ API, as well as a pre-trained ethnicity detector in the DeepFace algorithm (Serengil & Ozpinar, 2020). I removed individuals whose mode-apparent gender from all facial images did not fit their reported gender category. This is because there were some individuals who self-identified as the opposite gender. Like Wang and Kosinski (2018), I removed individuals whose mode-apparent ethnicity was not Caucasian to retain only Caucasians in the sample. I also excluded individuals whose average-apparent age detected from their photographs was not within the 18 to 40 age range. To ensure that the resulting gender and ethnicity in the dataset were accurate, I randomly drew 100 individuals from each gender-sexual-orientation category (a total of 400 images). All genders were perfectly classified. Only two persons' ethnicities[7] out of the 400 might be incorrectly classified (99.5% accuracy).

Machine learning research recommends balancing the training data to prevent bias towards the majority group (Susan & Kumar, 2021). Thus, I matched the sample size, age, and number of images of the sub-samples using an automated process. The process, conducted separately for each gender, paired every person from a sexual orientation group with a person from the other group by age and number of images. The resulting number of gay versus heterosexual people, as well as the number of

[7] One heterosexual man who appeared to be multiracial and one heterosexual man who appeared to be Latino were considered as misclassified. Note that this accuracy rate is high because strict inclusion criteria were used, i.e., included only if the mode ethnicity classification from all facial images of each person was Caucasian. Note also that Latinos were not considered as Caucasian by the DeepFace algorithm and by my manual accuracy check even when some might have identified themselves as Caucasian. Latinos were not included to avoid ambiguity.

their facial images, were completely balanced. The final sample contained 10,162 facial images of 5,124 men (50% gay and 50% heterosexual) and 21,600 facial images of 10,340 women (50% gay and 50% heterosexual). A breakdown of the final sample and age distributions is reported in Table 1.

Finally, I cropped and aligned all facial images using the bounding box provided by Face++ API. All resulting facial images in the dataset contained facial positions that matched exactly those reported in Figure 4 of Wang & Kosinski (2018). All images were resized to 224 by 224 pixels as required by the facial recognition algorithm used to classify sexual orientation.

----- INSERT TABLE 1 ABOUT HERE -----

Self-Presentational Attributes. A total of 12 self-presentational facial attributes was extracted from Face++ API to test whether there was a significant difference in these features according to sexual orientation (Study 1a), and whether this difference contributed to classification of sexual orientation (Study 1b). All attributes except head pose angles were measured in probabilities, i.e., how likely the face found in an image displayed a certain attribute. All attributes were standardized to a range of zero to one for easier comparisons. As participants varied on the number of facial images, a different number of sets of attributes was produced for each participant. The within-person reliability of these attributes was moderate (see Table 1 of Appendix), so within-person attribute sets were averaged so that each participant only had one set of facial attributes.

All six facial expression scores, such as happy and neutral, were included. These scores measured the probability that the face in each image displayed a certain facial expression. Research has shown that head pose angles are related to different facial perceptions and emotional expressivity (Barrett et al., 2019; Nicholls et al., 2002; Witkower & Tracy, 2019). Thus, head pose angles consisting of the absolute roll, yaw, and pitch angles were included. Research found that gay individuals reported

being more likely to wear glasses than heterosexual individuals (Agüera y Arcas et al., 2018; Wang & Kosinski, 2018). Thus, the probability of wearing glasses was included. Finally, eye status, the probability of eyes being open, and the probability of a smile being present in the image, were included.

Transparency and Openness. Data include sensitive personal information, thus would not be disclosed. The code is available at https://osf.io/q39py/. All materials are included in the main text. There are no additional materials to disclose. The design and analysis of this study were not pre-registered.

Results

Table 2 reports the mean, confidence intervals, and statistical tests of facial attributes. Figure 1 shows the differences in means ranked from positive to negative. All results were reported separately for women and men.

----- INSERT TABLE 2 ABOUT HERE -----

As indicated in Table 2, out of the 12 attributes tested in this study, 10 were significantly different across sexual orientations for women and six were different for men (p's < .05). As shown by Figure 1, a large difference was observed in the likelihood of individuals wearing glasses in facial images. Consistent across both women and men, gay people on average were more likely to upload photographs of themselves wearing glasses compared to heterosexual people in the sample. This is in line with the survey findings in Agüera y Arcas et al. (2018) that gay people on average preferred wearing glasses, as well as the aesthetics of wearing glasses, compared to using contact lenses or not wearing glasses. This is also consistent with the average faces reported in Figure 4 of Wang and Kosinski (2018).

----- INSERT FIGURE 1 ABOUT HERE -----

Discussion

In line with previous findings in the impression management and facial perception literatures, these results demonstrate significant differences in how gay and heterosexual people presented themselves in facial images. My study employed only a few facial attributes from the Face++ API, and found that most attributes differed across sexual orientations. Women on average demonstrated greater difference across sexual orientations, as seen by the larger effect sizes compared to men. Heterosexual women were more likely than lesbians to display facial actions such as turning the head sideways and smiling toward the camera, a facial expression resembling coyness that serves relationship-building functions and helps displayers to connect with observers (Keltner & Haidt, 1999; Reddy, 2000). This display aligns with previous research in the context of online dating (Todorov & Porter, 2014). Note that while impression management research posits that self-presentational motifs are typically high in dating contexts (Hancock & Toma, 2009; Leary et al., 1994; Tong et al., 2020), it is difficult to tell whether the self-presentational styles observed here were intentional or unintentional, and this question is beyond the scope of the study.

Study 1b

Study 1b aimed to examine the extent to which people's sexual orientation could be detected from their self-presentational facial attributes. I trained a logistic regression model using 20-fold cross-validation. If sexual orientation could be classified at rates above random chance, self-presentation was likely to have contributed to sexual orientation.

Methods

Self-Presentational Attributes. Study 1b employed the same self-presentational facial attributes extracted using the Face++ API from the dataset of facial images as Study 1a. Like Study 1a, whenever I had multiple images for the same individual, I averaged the attributes across all images.

Sexual Orientation Classifier. I trained a logistic regression model using 20-fold cross-validation. In each fold, the participants were split into 20 equal parts; 19 parts were used to train the logistic regression while the remaining part was used to classify the results. This process ensured that I never used the same data to train and classify the outcome.

Results

I report the area under the receiver operating characteristic curve (AUC) as a measure of the classification power. AUC is defined as the likelihood that when presented with two images, one from a gay person and one from a heterosexual person, the model would assign the gay person a higher likelihood of being gay than the heterosexual person. I also report the confidence intervals estimated using the DeLong method, a general practice in machine learning, deep learning, and facial recognition (DeLong et al., 1988). I used AUC because it is an evaluation metric widely employed in machine learning research, and was used in previous studies on this topic (Kosinski, 2021; Wang & Kosinski, 2018). I report other common evaluation metrics in Table 3. All results are reported separately for women and men.

As shown in Figure 2, classification power using all 12 self-presentational facial attributes extracted using Face++ equaled on average AUC = .609 (95% CI = [.598, .620]) for women and AUC = .551 (95% CI = [.536, .567]) for men. The two most predictive attributes were happiness expression and smiling. Happiness expression afforded AUC = .572 (95% CI = [.561, .583]) for women and AUC = .533 (95% CI = [.517, .549]) for men. Smiling afforded AUC = .576 (95% CI = [.565, .587]) for women and AUC = .544 (95% CI = [.529, .560]) for men. These results demonstrate that self-presentational facial attributes extracted using Face++, to some extent, contributed to classification of sexual orientation.

----- INSERT FIGURE 2 ABOUT HERE -----

----- INSERT TABLE 3 ABOUT HERE -----

Study 1c

The next step was to find out whether there are other self-presentational factors (i.e., the background) that revealed people's sexual orientation, and directly test their influence on the sexual orientation classification algorithm. I first replicated the algorithm following the exact procedures reported in Wang and Kosinski (2018). Note that both my study and the previous one relied on an off-the-shelf facial recognition algorithm. In other words, no deep neural network training was done. This was intentional because if an algorithm totally unrelated to sexual orientation classification could potentially be repurposed to classify sexual orientation, it would suggest a serious risk of privacy loss.

I then tested whether the image background contributed to sexual orientation classification by blocking out the facial portions of the images. Making deliberate adjustments or modifications to an image is called image augmentation in computer vision (Shorten & Khoshgoftaar, 2019; Zeiler & Fergus, 2013). In this case, I employed an augmentation technique called masking. If highly masked images, where only background information is retained, could classify sexual orientation at levels significantly higher than chance, it means that people of different sexual orientation are presenting themselves by choosing different image backgrounds on online dating websites.

Another goal of this study was to evaluate the limit of privacy protection. Masks are often worn physically in some parts of the world, and used digitally to block people's identities and to prevent the detection of certain private information (Matzner et al., 2016; Zhang et al., 2014). Thus, augmenting the image by masking facial portions aimed to test whether this privacy-protection strategy might prevent the loss of private information. If masks were effective, sexual orientation classification using masked images should drop to chance level. If not, it would imply a serious threat to privacy.

Methods

Facial Images. Study 1c employed the same dataset of facial images as Study 1a and 1b. However, instead of relying on the self-presentational attributes extracted using the Face++ API, I directly employed the preprocessed facial images.

Sexual Orientation Classifier. I replicated step-by-step procedures employed in Wang and Kosinski (2018) in extracting the deep neural network features, which were later used to identify individuals' sexual orientation. Specifically, I extracted the 4,096 scores for each facial image using the facial recognition algorithm, VGG-Face (Parkhi et al., 2015). Next, the 4,096 scores were reduced to 500 dimensions using the singular value decomposition, a technique like principal component analysis. Finally, the 500 dimension scores were passed through a logistic regression model, with L1-penalty of 1, to generate the binary sexual orientation classification of gay versus heterosexual. Note that no deep neural network was trained to classify individuals' sexual orientation; the network was merely used to convert images into 4,096 scores, and then a logistic model was used to guess people's sexual orientation from these scores.

I employed 20-fold cross-validation in training the classifier, which combined the singular value decomposition and logistic regression model. One concern might be why the data were not split using more common methods in machine learning such as the hold-out validation. In hold-out validation, data are split into training and testing sets only once. Since the sample size here is small compared to other machine learning studies (e.g., $n > 1$ million), the hold-out or unseen set might not be representative enough of the underlying distribution of the dataset, and thus have small power. For example, for men, the unseen test set would contain only 256 individuals if a 19:1 split was applied only one time. Repeating the validation 20 times using different combinations of the data would theoretically maximize the power of the study.

Another note is that the dataset contained multiple images for most individuals (see Table 1). Thus, all images of the same individual were assigned to one and only one cross-validation partition. In other words, if an individual has multiple facial images, the images never appeared in both the training and testing partitions in any cross-validation folds. More details on how this is programmatically implemented is shown in https://osf.io/q39py/. The resulting sample size by cross-validation fold, train-test partition, and number of images is reported in Table 2 of the Appendix.

A related question is why the data were not split into training, validation, and testing sets.[8] This is because, as mentioned earlier, the study employed a pre-trained algorithm (VGG-Face), and there was no hyper-parameter tuning or no model selection for the singular value decomposition and logistic regression models. Therefore, splitting the data further into training-validation sets was unnecessary (for a review on validation and model-selection methods, see Raschka, 2020).

Image Augmentations. To test the influence of image background on the sexual orientation algorithm, I applied a rectangular mask to the center of each facial image. The entire dataset of masked images was then used to classify sexual orientations using the same 20-fold cross-validated model pipeline described above. A total of 29 augmentations using masks of increasing size was conducted. In the first augmentation, 3.3% (100% divided by 30 augmentations) of the image was masked in terms of the image area and a classification score was recorded. In the last augmentation, 96.7% of the image was masked. Only a very tiny border, made up of 3.3% of the entire area, remained on the image. Additionally, I conducted an augmentation using a mask that covered 100% of the image to produce a random classification. When the image was entirely masked, the algorithm generated a random classification power. In this case, I verified that the VGG-Face algorithm produced facial scores of 4,096

[8] Typically, machine learning studies split the data into training and testing (i.e., unseen) sets. The training set would be further split into training and validation sets. The training sets would be used to train algorithms with different hyper-parameter settings. The validation sets would then be used to evaluate the models. After a best model was selected, the testing set would be used to evaluate the generalizability.

zeros. These scores produced zero-only predictions, which finally resulted in AUC = 0.500 (95% CI = [0.50, 0.50]). The first row of Figure 1 in the Appendix provides a few examples of the image augmentations.

Results

Sexual Orientation Classification. I first report results about the extent to which people's sexual orientation can be detected from their images, parallel to the main results of Wang and Kosinski (2018). The AUC results and their confidence intervals are reported in Figure 3. Unlike Study 1a and 1b, I report AUC results separately according to the number of facial images to replicate the format reported in Figure 2 of Wang and Kosinski (2018). I also report the average AUC score, which was calculated by averaging the classification scores across facial images within each participant. Corresponding confusion matrices that indicate the accuracy of the model's classification are reported in Figure 4. Other common evaluation metrics are reported in Table 4.

The model's average AUC was .702 (95% CI = [.692, .712]) for women and .662 (95% CI = [.647, .677]) for men. AUC increased for both women and men when more facial images were used per person. For five facial images, the AUC increased to .732 (95% CI = [.691, .774]) for women and .797 (95% CI = [.736, .858]) for men. This AUC was similar to those found by Wang and Kosinski (2018) for women, however for men, this AUC was lower than Wang and Kosinski's AUC of around .81 when one image was used per person. Overall, the above results confirmed that it is possible to detect sexual orientation from images posted on the dating website. Next, I examined how masking the facial images affects the model's classification rate to assess the likely impact of a self-presentational feature on sexual orientation detection.

----- INSERT FIGURE 3 ABOUT HERE -----

----- INSERT FIGURE 4 ABOUT HERE -----

----- INSERT TABLE 4 ABOUT HERE -----

Image Augmentations. Figure 5 shows the AUC results when different proportions of the facial image have been masked. For ease of interpretation and brevity in the main text, I report the AUC scores averaging across the predictions from the varying levels of facial images per participant. Table 3 of the Appendix reports the AUC scores and significance tests against random classifications. For compatibility with previous research, I also report AUC scores generated using only one image per person in Table 4 of the Appendix. Tables 5 and 6 of the Appendix report other common evaluation metrics using averaged predictions or only one prediction respectively. All results are reported separately for women and men.

Of the 29 degrees of masking, no AUC dropped to chance level (AUC = 0.50) for both women and men. All AUCs were significantly higher than random chance ($p < .001$). As seen in Figure 5, AUC scores started to drop when masks were applied to each facial image. Nevertheless, these scores remained above random chance throughout. When the face was effectively masked in each facial image (mask area = 50% of the entire image), the AUC scores for gay vs. heterosexual people translated to Cohen's $d = 2.295$ for women and $d = 1.631$ for men, which are huge effect sizes (Sawilowsky, 2009). As I applied larger masks, the AUC continued to degrade but never reached random-chance level. At the most extreme case, when 96.7% of the image was masked, AUC scores for both women and men were above random-chance level and translated into Cohen's $d = 0.760$ for women and $d = 1.072$ for men (on average large effect sizes).

----- INSERT FIGURE 5 ABOUT HERE -----

While AUC scores degraded as images were progressively masked, the degradation did not follow the same magnitude of degradation for image pixels. One interesting finding here relates to how AUC degradation differed by gender; AUC degraded more severely at the beginning for women compared to men. The degradation started to show signs of flattening but picked up again towards the end. However, the degradation pattern followed an almost curvilinear pattern for men, where magnitude of degradation started to drop towards the end. For women, the results suggest that the facial regions might have contributed more compared to background, and vice versa for men. This seems to be consistent with the findings in Study 1a, in which women on average displayed larger differences in facial attributes between sexual orientations.

Discussion

Study 1c demonstrated that sexual orientation classification was possible using the dataset I collected, replicating the results in Wang and Kosinski (2018). Study 1c also found that the image background or pixels at the image border contributed to sexual orientation classification. These findings were alarming because in the most extreme case, masked images contained only 3.3% of the original pixels. Yet, these pixels could generate above-random-chance classifications. When interpreted together with the findings in Study 1a and 1b, these findings suggest it is possible that gay vs. heterosexual people chose to upload photographs using different self-presentational facial attributes as well as different backgrounds on the online dating website. The differences in turn allowed the model to classify their sexual orientation using these self-presentational features. Importantly, these findings also warn about a privacy loss that was previously underestimated, because masking the face might not effectively prevent the exposure of sexual orientation in facial images, contrary to previous research in privacy protection (Zhang et al., 2014).

Study 2a

The goal of this study was to assess whether the overall brightness of the image was yet another self-presentational variable that could reveal people's sexual orientation. Previous research found that gay men's average faces had brighter skin tones (Wang and Kosinski, 2018). However, it could be possible that skin tones were related to the overall brightness or illumination of the images. To test this, I separated the image into facial region (50% of the image area) and image background (50% of the image area). I extracted the average brightness of these regions, and examined the brightness by regions and sexual orientation. If the brightness of the facial regions varied according to sexual orientation in the same direction and magnitude with background, it would be likely that the appearance of skin tone was related to the overall brightness or illumination of the image. In this case, the difference in the general brightness of the image would suggest a possible form of self-presentation, in which people in the sample chose to present themselves in brighter or darker images or locations that were illuminated differently (e.g., indoors vs. outdoors) according to their sexual orientation.

Methods

Facial Images. I employed the same dataset of facial images as Studies 1a, 1b, and 1c.

Image Regions. To separate the facial image into different regions, I designed a fixed mask using the average facial landmarks extracted from each facial image. Since all faces were aligned at the same location, a fixed mask would be reasonably effective in covering all faces (or backgrounds) in the images. Figure 6 shows an example of the masks used to separate the regions. The size and shape of the mask were optimized such that the mask only occupied 50% of the entire image, at the same time covering a slightly larger area of the average facial region (see average facial landmarks in Figure 6). I also ensured that the mask would cover all the skin area including the neck. To extract the background regions, I blocked the facial region using the mask (see Figure 6, left panel). To extract the facial regions, I inverted the mask and blocked the background regions (see Figure 6, right panel). I applied the standard and

inverted mask onto every facial image to produce two masked images for each original facial image. As the number of facial images per participant is different, I created a composite image for each person.

----- INSERT FIGURE 6 ABOUT HERE -----

Image Brightness. Image brightness was measured by the average value of all pixels in the image for each color channel. The higher the value of the pixel, the brighter the image. Since pixels are valued from 0 to 255, I normalized them to a range of 0 to 1. I extracted two within-person brightness measures, i.e., the brightness of the background and that of the facial region for each participant.

Results

Figure 7 plots the image brightness by image regions and sexual orientation separately for each gender. Table 7 in the Appendix reports the statistical tests on the differences in brightness for each color channel by image regions and sexual orientations. Note that the general trend in how brightness varied was similar across each color channel (red, green, and blue). This meant that the difference was indeed due to brightness and not hue (see Table 7 in Appendix). Looking at Figure 7, the facial regions were across the board brighter than the background, which confirms that the masks correctly separated the facial regions and the background. A significant difference in image brightness was observed between gay and heterosexual people in the sample for both genders and facial regions.

----- INSERT FIGURE 7 ABOUT HERE -----

On average, images were significantly darker for lesbian women compared to heterosexual women ($p < .001$), whereas images were significantly brighter for gay men compared to heterosexual men ($p < .001$). Comparing the regions, image background was on average darker than the facial region,

but the magnitude and direction of the variations were similar. In other words, if facial regions were on average darker for a given sexual orientation, the background region would also be darker for that sexual orientation compared to the other group. This supports the hypothesis that skin tone differences across sexual orientations was related to differences in the overall brightness of the image and not just the facial regions. This suggests a form of self-presentation, in which people preferred different image brightness or illumination according to their sexual orientations.

I further conducted a mixed-design ANOVA separately for each gender. There was a significant main effect of masked regions on brightness differences for both women ($F(1, 10338) = 4711.06$, $p = < .001$, $\eta p2 = .313$) and men ($F(1, 5122) = 199.14$, $p = < .001$, $\eta p2 = .037$). There was also a significant main effect of sexual orientation on brightness differences for both women ($F(1, 10338) = 19.32$, $p = < .001$, $\eta p2 = .002$) and men ($F(1, 5122) = 42.22$, $p = < .001$, $\eta p2 = .008$). However, I did not observe any significant interaction between masked regions and sexual orientation in terms of brightness difference for both women ($F(1, 10338) = 2.64$, $p = .104$, $\eta p2 = < .001$) and men ($F(1, 5122) = 0.037$, $p = .847$, $\eta p2 = < .001$).

Discussion

Study 2a found that brightness in the facial region varied together with background region according to sexual orientation. For example, gay men on average uploaded facial images that are brighter in both the facial and background region compared to heterosexual men. The brighter skin tone observed in the previous study (Wang and Kosinski, 2018) for gay men might have been due to gay men taking and uploading generally brighter or more illuminated facial images to the dating website. In other words, the brighter images for gay men might have made their skin appear brighter compared to heterosexual men. Overall, this difference in image brightness suggests a coherent story of how gay and heterosexual people in the sample chose to upload different facial images according to their self-presentational preferences.

The difference in illumination of facial images across sexual orientation and gender groups might be attributed to many reasons. One interpretation might be the use of camera flash lights. However, flash lights should be directed to the face and should not affect the overall brightness of the image. Another possible interpretation might be that people used different photo-editing processes, such as applying image filters (Haferkamp et al., 2012; Hancock & Toma, 2009; Ota & Nakano, 2021). A third interpretation might come from theory in person-environment transactions, which posits that people actively choose their daily environments according to their dispositions (Matz & Harari, 2021; Wrzus et al., 2016). It would be possible that for people with brighter images, the images were taken outdoors. This interpretation might also help to explain why image background in Study 1c contributed to sexual orientation classifications for men. That said, it might be possible that a combination of factors, i.e., flash lights, filters, locations, and even skin tones, contributed to the observed differences, and that these factors played out differently depending on the gender.

Study 2b

Study 2b tested whether differences in image brightness played a significant role in classification of sexual orientation. To do so, Study 2b employed an image augmentation technique: blurring. I blurred the images progressively and tested the classification power of blurred images using the sexual orientation classifier. When images were completely blurred, only brightness differences remained. This helped me measure the contribution of brightness on classification of sexual orientation. If three values that measured the image brightness for each color channel can classify sexual orientation, the classifications are likely attributed to self-presentation.

Importantly, Study 2b was also designed to evaluate the limit of privacy protection in sexual orientation classification. Image blurring is a frequently used strategy to de-identify facial images and to protect the privacy of individuals (Li et al., 2017). Augmenting the image by blurring aims to test whether this privacy-protection strategy may prevent the loss of private information. If blurring were

effective, sexual orientation classification using blurred images should drop to chance level. If not, it would reveal a threat to privacy and a limit to privacy protection.

Methods

Facial Images. I employed the same dataset of facial images as Studies 1a, 1b, 1c, and 2a.

Image Augmentations. The second row of Figure 1 in the Appendix provides a few examples of the image augmentations. I blurred each image by downsizing it to a target width and height. The downsizing algorithm interpolated every new pixel using the average values of its respective source pixels.[9] As the algorithm employed in this study requires images of size 224 by 224 pixels, I enlarged every image back to its original size after it was downsized.

The entire dataset of blurred images was used to classify sexual orientations following the same 20-fold cross-validated model pipeline described in Study 1c. A total of 29 augmentations using target width of decreasing sizes was conducted. Figure 2 in the Appendix provides a plot of the target width by augmentation number. In the first augmentation, a target width of 112 pixels was used (50% of the original image width) and a classification score was produced. This step was repeated until, in the last augmentation, the image was downsized to only one pixel. In other words, it was completely blurred.

To evaluate these classifications, I compared them to random classifications. Unlike Study 1c, random classifications were generated using the downsized average image of the entire dataset (separately for each gender). As expected, blurred average image of the entire dataset generated a random classification score of AUC = 0.500 (95% CI = [0.50, 0.50])[10] for each gender.

---

[9] For example, if a 224 by 224 image were downsized to one pixel, the resulting image would contain the average pixel values of the original image.

[10] In this case, VGG-Face algorithm produced random facial scores (mostly zeros), but when they entered the 20-fold cross-validated logistic regression, they yielded zero-only classifications like Study 1c.

Results

Figure 8 shows the average AUC results with confidence intervals. Table 8 of the Appendix Materials reports the average AUC results and significance test against random classifications. Table 9 of the Appendix reports AUC scores generated using one image per person. Tables 10 and 11 of the Appendix report other common evaluation metrics using averaged predictions or only one prediction respectively. All results are reported separately for women and men.

Of the 29 degrees of blurring, no AUC dropped to chance level (i.e., AUC = 0.50) for both women and men. All AUCs were significantly higher than random chance ($p < .001$). As seen in Figure 8, slightly blurring the facial images did not lead to any decrease in AUC scores for both women and men (i.e., for the first four augmentations). In fact, AUC scores even increased for men when facial images were blurred to 44 by 44 pixels. This could be due to blurring eliminating noisy pixels, thus allowing the algorithm to "focus" better on the useful pixels.

As I increased the degree of blurring, the AUC scores remained high for both women and men but started to decrease until the midpoint of the augmentations. At this point, the facial features started to become too blurred to be recognized, as every facial image was downsized to only 16 by 16 pixels. Despite this, the AUC scores remained high at AUC = .62 for women and AUC = .59 for men. These scores translated to Cohen's d = 2.36 for women and d = 1.50 for men, which are on average huge effect sizes. As the blurring went along further, AUC started to degrade more for women but remained almost constant for men. Finally, at the most extreme case, i.e., when images were downsized to one pixel, AUC scores for both women and men were above chance level and translated into Cohen's d = 0.55 for women and d = 1.00 for men (i.e., large effect sizes).

----- INSERT FIGURE 8 ABOUT HERE -----

Looking at the overall AUC degradation pattern, some obvious differences across genders were observed, as in Study 1c. AUC scores degraded as pixels were progressively blurred, but the decrease in scores did not follow the same magnitude of degradation in image quality. AUC degradation seemed to follow a linear pattern for women but a slightly curvilinear or even somewhat random pattern for men. Intriguingly, for men, the degradation started to show signs of increase beyond the point where facial features became unrecognizable (when downsized to 6 by 6 pixels) compared to earlier augmentations. It is interesting that this general trend for both genders was somewhat consistent with Study 1c, in which the AUC degradation was slightly more severe for women but not so for men.

Discussion

These results, when interpreted together with those in Study 1c, might provide a consistent story about how men chose photographs with different backgrounds and image lighting (possibly due to different locations or ambient environments). Finally, like Study 1c, these results demonstrated a serious threat to privacy. This is reflected in the most extreme case, in which the only information necessary for classifications could be as few as three numbers representing the brightness values of each color channel.

General Discussion

This research examined whether gay and heterosexual people presented themselves differently in facial images that were uploaded to a dating website, and whether these differences contributed to classification of sexual orientation. Study 1a showed that self-presentational facial attributes extracted from photographs differed significantly across sexual orientations. Study 1b showed that these differences contributed to classification of sexual orientation. Study 1c replicated the sexual orientation classifier in Wang and Kosinski (2018) and examined the contribution of image background on sexual orientation classification. The results showed that the classifier was influenced by the image background and that even a 97% masked facial image could classify sexual orientation at rates higher than random

chance. Several robustness checks revealed that this pattern was not due to artefacts in the algorithm. Study 2a examined another self-presentational feature: image brightness. Results showed that the difference in skin tone across sexual orientation was due to overall brightness or illumination of the image. Given that brightness varied across sexual orientation, Study 2b tested whether highly blurred images, where only brightness information remained, would classify sexual orientation. As expected, results showed that image brightness contributed to classification of sexual orientation, particularly for men. These results demonstrate that people presented themselves differently in facial images, classification of sexual orientation was influenced by these differences, and the risk of privacy loss might be higher because de-identification strategies were ineffective.

Theoretical Contribution. If self-presentation mattered for classification of sexual orientation, what about facial morphology? Wang and Kosinski (2018) acknowledged the contribution of both morphology and self-presentation but framed their study using the Prenatal Hormonal Theory. For example, they wrote, “Some of the differences between gay and heterosexual individuals, such as the shape of the nose or jaw, are most likely driven by developmental factors” (p. 254). My study aimed to provide an alternate narrative by showing that self-presentation makes a significant contribution to sexual orientation classification. For example, 12 self-presentational facial attributes could classify sexual orientation at AUC scores almost comparable to when the entire facial image was used (4,096 VGG-scores) for women.

One question that might arise is which of these influences, facial morphology or self-presentation, contributed most to classification of sexual orientation. There is ample evidence in facial perception research pointing to confounds in facial images, and how these confounds might render photographs unreliable measures of facial identity, appearance, or morphology (Jenkins & Burton, 2011; Jenkins et al., 2011). For example, Jenkins et al. (2011) found that within-person variability in photographs is sometimes even larger than between-person variability, which caused unfamiliar raters

to misidentify faces in the photographs. Noyes and Jenkins (2017) found that camera-to-person distance can also moderate the person's face shape in photographs, leading to different levels of face recognition by humans. Consistent with this line of research, my study would argue that my sample, i.e., dating photographs, is not fit to answer morphological related questions. It would be likely that facial morphology and self-presentation both contributed to classification of sexual orientation, and there might be a significant interaction between the two.

Future research interested in the morphological contribution of sexual orientation classification should consider using standardized facial images that are taken in controlled conditions (such as the lab). Future studies should also control for the influence of important self-presentational features such as the presence of glasses and image backgrounds. In particular, social cognition researchers typically use stimuli, in which all non-facial parts of the face (including hair) are occluded (e.g., see Oosterhof & Todorov 2008, Figure 1). Researchers can use such controlled, standardized stimuli, in which there is no room for self-presentation, to determine whether sexual orientation can be detected from morphology.

These arguments could be easily dismissed on grounds that advanced facial recognition algorithms are less affected by self-presentational features (Phillips, 2017; Prasad et al., 2020; Taigman et al., 2015). For example, the VGG-Face algorithm employed in this study was developed using 2.6 million facial images of 2,622 unique individuals (Parkhi et al., 2015; Phillips, 2017; Prasad et al., 2020). Given that, on average, 1,000 images per person were fed into the algorithm, researchers believed it would start to recognize high-level facial patterns, such as the contour of the jaw or the shape of the nose and ignore confounding features such as the image background. However, consider the following illustration. Diane is an extrovert. She enjoys outdoor activities and likes taking selfies of herself outdoors. Mary is an introvert. She likes to stay at home and rarely takes photographs of herself. When they uploaded photographs to their social media platforms, Diane consistently chose photographs of her outdoors while Mary chose only a picture of her in her bedroom. Therefore, their personality traits, such

as whether they are extroverted, might be consistently reflected in their self-presentational behaviors, e.g., lighting of their photographs, skin tan, and image background, but not so much in their facial morphology or appearance.

My findings suggest that facial recognition algorithms might not have eliminated self-presentation features. While facial recognition algorithms perform well under different lighting conditions (Phillips, 2017), it does not mean that differences in lighting or brightness, a form of self-presentational feature as seen in Studies 2a and 2b, are disregarded by the algorithm. The findings in Studies 2a and 2b suggest that when the algorithm was applied to sexual orientation classification (from facial recognition), it started to rely on such features. Thus, future research interested in facial morphology should consider methods in other areas of machine learning. For example, 3D face reconstructions using multiple facial images of the same person (Gecer et al., 2019; Tran et al., 2018) might produce more accurate morphological models, might be less affected by self-presentational features, and might more robustly model the relationship between morphology and behavioral outcomes (Kittler et al., 2016).

In terms of impression management, the findings in this research are consistent with the idea that people present themselves differently based on their values, goals, identity, and social contexts (Goffman, 1959; Leary & Kowalski, 1990; Schlenker, 2012). My study found that people in the sample systematically uploaded different photographs to the dating website according to their sexual orientation and gender. For example, men on average uploaded photographs that differed significantly in image background and brightness across sexual orientations. As elucidated, while self-presentational differences were observed, impression management research suggests that self-presentations can be either conscious or subconscious (Leary & Kowalski, 1990). It is beyond the scope of my study to further examine the motivations.

The findings here also contribute to facial perception research. The literature has established a vast framework showing how morphological and self-presentational differences influence social evaluations (Oosterhof & Todorov, 2008, 2009; Todorov, 2008; Todorov et al., 2008). However, it seems there is a scarcity of research examining how image backgrounds or lighting in naturalistic facial images affect perceptions of faces. Future facial perception research can explore whether image properties affect the impressions of people, and how properties interact with morphologies or other self-presentational styles in forming facial perceptions. The findings here also seem to connect to the burgeoning literature around the biasing nature of faces (Olivola et al., 2014; Olivola & Todorov, 2010; Todorov et al., 2015). If perceptions of faces can be biased, how do different self-presentations identified in this study affect or improve the accuracy of these perceptions? Future studies at the intersection of impression management and facial perception research could perhaps explore these questions further.

Finally, my study might relate to research in person-environment transactions (Matz & Harari, 2021; Wrzus et al., 2016). When correlating naturalistic facial images with behaviors, it is often impossible to eliminate the influence of image background. In fact, the findings here might demonstrate what was shown in Torralba et al. (2008), that scenes can be consistently predicted using tiny images. I did not directly test whether the difference in background and lighting was caused by ambient environmental conditions (or location), but if these conditions influenced the results here, it would open new doors to research. Perhaps future research can evaluate whether people of different characteristics or personality traits systematically take selfies in different locations, whether these locations are correlated with other self-presentational styles, and how these locations further influence people's psychological states.

Practical Implication. Most importantly, my research revealed a danger to privacy that was underestimated in previous research. The results here suggest that a small number of pixels at the

image border or even one pixel representing the average brightness of the image can expose sexual orientation information. Therefore, masking or blurring the face in naturalistic facial images does not completely prevent the loss of this sensitive information. One counter-argument is that classification of sexual orientation using self-presentational differences might be easier in the context of dating but more difficult in other contexts. However, previous research might suggest that some self-presentational behaviors are habitual or consistent, thus do translate to other contexts (Agüera y Arcas et al., 2018). For example, gay individuals reported that they wore glasses more (and preferred the aesthetics of doing so) compared to heterosexual individuals. Therefore, "non-transient" self-presentational factors might consistently reflect in other contexts, such as job interviews, and expose sensitive information.

We live in a day when we are closely connected to the digital world and acculturated to expressing ourselves through online platforms. However, companies such as Clearview AI are taking advantage of this trend by amassing huge databases of facial images without our consent, and training facial recognition algorithms to expose our sensitive information in the name of law enforcement (Rezende, 2020). While research suggests that consumers can employ do-it-yourself methods to protect themselves (Li et al., 2017; Matzner et al., 2016; Shan et al., 2020; Zhang et al., 2014), the findings here indicate that these methods might not be as effective as researchers imagined. If governments and companies do not take the necessary steps to address this matter, the alternative to privacy protection would be people uploading fewer photographs and engaging in less online activity. However, this would entail a tradeoff between privacy and freedom of expression, one that would be detrimental to our society. Overall, based on the results in this study, I recommend against shifting the burden of privacy protection to consumers. Fortunately, some governments have started to limit the use of certain facial recognition systems (Conger et al., 2019). They are also implementing stricter privacy protection laws

(Rothstein & Tovino, 2019). It is my hope that with this trend continuing, people will not face the privacy concerns addressed in this study.

Limitations. One question regarding Studies 1c and 2b might be whether the pre-trained VGG-Face algorithm assisted the classification when the images were masked or blurred. In other words, were there artefacts in the pipeline or algorithm? To answer this question, I correlated the VGG-scores produced by the image augmentations with the scores produced by the original images. I found that the correlations decreased as images were progressively masked. When the images were fully masked, VGG-Face produced zero-only scores. When fully blurred images of the dataset were entered, VGG-Face produced the same random scores (mostly zeros) for every image. These scores, when fed into the logistic regression model, all produced predictions of zero, resulting in AUC = 0.50 (95% CI = [0.5, 0.5]). These results show that the VGG-Face algorithm did not assist with the classifications and there were no artefacts in the pipeline. Additionally, I repeated the entire study by custom-training the VGG-Face algorithm (i.e., not using the pre-trained weights) using the augmented images for each of the 29 augmentations. This process produced even higher classification powers, which might indicate over-fitting. As an additional robustness check, I repeated the steps using different validation techniques. The results remained similar. These analyses confirmed that the classification powers reported here were indeed due to background or image brightness and not due to the “robustness” or artefacts of the VGG-Face algorithm.

A related issue might be why image augmentations were used to explain the algorithm in the first place. To elaborate, image augmentation is an attribution technique used in machine learning to increase interpretability (for other methods, see Biecek, 2018). This technique is not new and has been widely applied to many areas of machine learning (Shorten & Khoshgoftaar, 2019; Zeiler & Fergus, 2013). However, I deliberately chose image augmentations, i.e., masking or blurring, because they are straightforward and offer results that were easy to interpret. More importantly, the augmentations

addressed the research questions in this study. One question was whether the background mattered for classification of sexual orientation. By masking the facial regions of the image, Study 1c showed that the background contributed to classification of sexual orientation. Another question was whether image brightness contributed to classifications. Thus, Study 4 blurred the image so that only the brightness information was used to classify sexual orientation. Lastly, the most important research question was whether data-protection tools would help people protect their private information (Li et al., 2017; Matzner et al., 2016; Shan et al., 2020; Zhang et al., 2014). Image augmentations directly tested whether such strategies were effective. Nevertheless, future studies can employ different attribution techniques and interpret the results here in tandem.

A question regarding the classification results might be why AUC scores are generally lower in my study, especially for men, compared to Wang and Kosinski (2018), despite following that previous study closely in every procedure. First, the lower AUC scores might be due to the smaller sample. The total number of facial images in my final dataset is 31,762, whereas the total in Wang and Kosinski (2018) is 35,326. While my study aimed to collect a larger dataset, the preprocessing step pruned more samples compared to the previous study as participants were also matched on the number of facial images (see Table 1). Second, if the lower AUC scores were due to my reliance on a different sample (i.e., a different dating website), the fact that my study replicated the general trend in Wang and Kosinski (2018) shows that classification of sexual orientation is generalizable, providing evidence for even greater privacy concern.

A relevant related question might be whether the AUC scores in the most extreme augmentations were too low to warrant any privacy concerns. The augmentations were deliberately exaggerated to examine whether there was a baseline, at which AUC would drop to chance-level. In other words, it would be unlikely that images in the wild would contain augmentations that were as severe as those in the 29th step of the augmentations. Even if they did, it is highly likely that companies,

such as Clearview AI, possess more facial images (in the hundreds) of the same person. These images might be used to produce much higher classification rates.

Conclusion

This study examined three research questions, namely whether people presented themselves differently in naturalistic facial images according to their sexual orientations, how differences contributed to classification of sexual orientation, and the implications for privacy. The research found that differences in self-presentation across sexual orientation were significant and that these differences contributed to classifications. Image augmentations found that even when the face was masked or blurred, sexual orientation was classified at rates significantly higher than chance, signaling a risk to privacy that was previously underestimated. Given these findings, I argue that the burden of privacy protection should not be shifted to the consumers, but must be initiated by governments and companies.

Chapter 2

How Existing Biometric Privacy Acts Would Fail at Protecting People's Privacy in Self-Posted Facial Photographs

Introduction

In recent years, more evidence has emerged that facial recognition is violating people's privacy, threatening civil liberty, freedom, and safety in our society (Leong, 2019; Matz et al., 2020; Santow, 2020). A number of studies have demonstrated how facial recognition can be easily repurposed to reveal people's personal information from their online-posted facial photographs (Kosinski, 2021; D. Wang, 2021; Y. Wang & Kosinski, 2018). This information not only includes basic demographics (Ranjan et al., 2018), but also personal features such as sexual orientation (D. Wang, 2021; Y. Wang & Kosinski, 2018), political orientation (Kosinski, 2021), personality traits (Kachur et al., 2020), and emotional states (Schwartz, 2019). Using artificial intelligence technology, companies can discretely engage in inappropriate activities, such as changing people's purchasing behaviors (Matz et al., 2017), profiling people according to their race and psychological traits (Matz et al., 2020), and swaying public opinions and political views (Matz et al., 2017). The potential for facial recognition to be misused or even weaponized poses a significant threat to our society, a challenge that must be addressed urgently.

The Current Regulatory Environment

To tackle these concerns, many governments have started to implement different kinds of privacy acts (Almeida et al., 2021). One type is the consumer privacy act. For example, the California Consumer Protection Act of 2018 (CCPA) (*California Consumer Privacy Act (CCPA)*, 2018) requires companies to ask consumers for consent when processing their data, including biometric-related data, and to provide consumers the liberty to have their data deleted. Facial recognition would be regulated

because both the input and output of the system, i.e., consumers' facial photographs and biometric information, are within the scope (Helveston, 2018). However, this regulation is often criticized for the fact that companies are not required to publicly disclose how they engage in data-processing activities, and thus the burden of privacy falls on the shoulders of the consumers. Relying on consumers' discretion has proven ineffective from a psychological perspective, as published research illustrates how consumers often cannot comprehend disclosure documents or are misled by ambiguous framing of privacy disclosure terms (Acquisti et al., 2020).

On the other hand, many governments are considering a different form of regulation. This type is distinct from the above-mentioned as it targets only the output of facial recognition—biometric information. Furthermore, it requires companies to not only ask for consent, but also disclose how they are using the biometric data. A notable example is the Illinois Biometric Information Privacy Act of 2008 (BIPA), which specifies that any "retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry" that could be used to "identify an individual" is regulated under this statute (BIPA, 2008). Moreover, companies "must develop a written policy, made available to the public" before biometric information is obtained or processed (BIPA, 2008). Recently, a watershed moment was reached when a class action lawsuit against Facebook under BIPA violation was settled, amounting to potentially 650 million dollars of payout (Stempel, 2019). A summary of the current regulatory environment under biometric privacy regulations can be found in BCLP Law (2022).

Biometric Information vs Non-Biometric Information

One major assumption, which proved pivotal to the settlement of a number of court cases under existing biometric privacy acts (Stempel, 2019), is that privacy loss is a result of facial recognition's power to extract biometric-equivalent information from facial photographs posted online (Sundararajan & Woodard, 2018). This assumption is crucial because existing biometric privacy acts exclude facial photographs in their definition of biometric data (BIPA, 2008; *Norberg v. Shutterfly*, 2015). Therefore,

for the law to be effective at regulating facial recognition, it must assume that the system can recognize faces and extract biometric information (Q. Cao et al., 2018; Parkhi et al., 2015; Prasad et al., 2020; Sundararajan & Woodard, 2018; Taigman et al., 2015). This assumption can also be supported by an abundance of evidence in social psychology, evolutionary psychology, and behavioral sciences that demonstrates robust links between faces (particularly facial geometry) and behavioral tendencies (Oosterhof & Todorov, 2008a, 2008b; Stirrat & Perrett, 2010; Todorov et al., 2005). However, recent research demonstrates that this basic assumption might not be valid (Geirhos et al., 2020; Rudin, 2019), for it would be possible that facial recognition learns non-biometric information too (Agüera y Arcas et al., 2018; D. Wang, 2021), and harms privacy due to features in the model previously overlooked by policymakers and legal practitioners.

In the technical sense, biometric-equivalent information extracted using facial recognition systems is usually assumed to be the string of numbers at the end of the facial recognition model (see Figure 1a). This string of numbers is referred to as face-embeddings (Q. Cao et al., 2018; Parkhi et al., 2015; Prasad et al., 2020; Schroff et al., 2015). Like fingerprints, face-embeddings are unique to each individual because the numerical values in the embeddings would largely remain similar regardless of the facial photographs given to the facial recognition system (Phillips, 2017). VGGFace, for example, is a pre-trained facial recognition model that extracts a unique identifier of 4,096 numerical values for each individual (Parkhi et al., 2015). These numbers remain roughly identical across different pose, illumination, and age of the same person (Q. Cao et al., 2018; Parkhi et al., 2015; Phillips, 2017), and thus can be used to verify the identity of the person or distinguish the person from others.

----- INSERT FIGURE 1 ABOUT HERE -----

However, given that deep learning systems are largely black-boxes (Rudin, 2019), researchers are starting to question whether face-embeddings extracted from photographs consist of *only* biometric-equivalent information (D. Wang, 2021). The existence of non-biometric information in facial-embeddings would be possible because facial photographs posted online are usually voluntarily generated by people, thereby mingled with their behavioral tendencies (see Figure 1b). For example, psychological research has found that when people upload profile pictures to their social media accounts, they do not randomly upload their photographs (Todorov & Porter, 2014; White et al., 2017). Instead, they select photographs that they consider to best represent themselves (Agüera y Arcas et al., 2018; D. Wang, 2021). They also add specific modifications such as image filters, different illuminations or brightness, or crop in or out objects that do or do not symbolize their values (White et al., 2017). These subtle behavioral cues, more often than not, are consistent across photographs of the same person (Agüera y Arcas et al., 2018).

Thus, it would be highly likely that these behavioral features are also learned by the facial recognition model and would carry significant value in exposing people's privacy. If companies could leverage on behavioral data extracted from photographs, instead of biometrics, to expose people's sensitive information, the companies could still invade people's privacy without the restrictions of biometric privacy acts. Such data can also be combined with other digital traces (such as text data) to increase prediction accuracies (West, 2019). Since the scope of existing biometric privacy acts do not include behavioral data (BIPA, 2008), if the abovementioned is true, it would not be broad enough to fully protect people's privacy.

In this research, we demonstrate a case in which sensitive information can be easily extracted without the help of a facial recognition system and biometric-equivalent information. We repurposed a face, image, and scene recognition model into trait classifiers. We found that the image and scene recognition models were capable of predicting the human traits at levels comparable to the facial

recognition model. We further de-identified individuals in the photographs by segmenting and masking the person in every photograph. In all cases, non-biometric models predicted traits at levels either comparable to or better than facial recognition models. The findings in this research reveal how the use of non-biometric information could endanger privacy, thereby rendering the current scope of biometric privacy acts incomplete. We hope our results contribute to the urgent discourse surrounding privacy regulations.

Methods

Data and Sample. We obtained two publicly available datasets, CelebA-HQ (Guo et al., 2016; Karras et al., 2018) and UTK-Face (Z. Zhang et al., 2017). While there are many open-source datasets, we chose these two because photographs were labeled with attributes of interest. Moreover, these datasets contain raw, unprocessed, and unaligned photographs found in realistic online settings. The CelebA-HQ dataset consists of 30,000 high-quality photographs of 2,222 celebrities collected in-the-wild, annotated with gender and age labels (Guo et al., 2016; Karras et al., 2018). The UTK-Face dataset consists of 24,102 photographs, one per person, collected from the Internet, annotated with gender, age, and race labels (Z. Zhang et al., 2017).

As the number of prediction classes and images per person were somewhat different, we balanced the dataset by randomly drawing samples in each sub-category using a target sample size equivalent to but not more than the smallest sub-category. In the final sample, the number of images in each prediction class were perfectly balanced, as well as the number of images per person. Table 1a reports the size and distribution of the final sample for UTK-Face and Table 1b reports that for CelebA-HQ.

----- INSERT TABLE 1 ABOUT HERE -----

Pre-trained Models. All embedding features (except the baseline) were extracted using deep neural network models. The first model was a facial recognition model called VGGFace (Parkhi et al., 2015), trained on 2.6 million naturalistic facial images of 2,622 individuals. A number of studies have demonstrated how this model can be repurposed to detect sensitive traits, such as sexual orientation, at high accuracy rates (D. Wang, 2021; Y. Wang & Kosinski, 2018). Thus, we believe this pre-trained model would be appropriate to demonstrate our proposed effect. We replicated the preprocessing pipeline in previous research and took the performance of this model as the benchmark.

In comparison, we obtained two off-the-shelf image and scene recognition models, namely ImageNet and Places-365. Both models were repurposed to predict human traits, far from their original intentions. The image recognition model was trained using a small subset (1.3 million) of the complete ImageNet dataset (Deng et al., 2009). These images include labels such as dog, cat, cup, and baseball. Similarly, the scene recognition model was trained on a small subset (1.8 million) of the complete Places-365 dataset (Zhou et al., 2018). These images include labels such as bathroom, bedroom, and warehouse.

Image Preprocessing and Training Pipelines. To demonstrate a case in which non-biometric information in photographs posted online can reveal sensitive information, we reconditioned a face, image, and scene recognition model to predict three demographic traits: gender, age, and race. Note that while it is possible to demonstrate the effect using intimate human traits such as sexual orientation, we avoided doing so due to ethical concerns—we believe it unnecessary to commit a crime to demonstrate the criminality of an action. To compensate for the robustness of the research, we conducted conservative tests by de-identifying the person in each photograph and evaluating the performance using the same models. We replicated the results on two publicly available datasets, as well as made our code available for future research (see Code Availability).

Figure 2a illustrates the preprocessing and feature-extraction pipeline. For each photograph, we obtained a total of 10 sets of embedding scores. In the first two, a facial recognition model was applied on the aligned and unaligned photograph to extract the embedding scores (see upper half of Figure 2a). For the next two sets, we used an image and scene recognition model. As baseline, we included an analysis without the reliance of any deep neural network model. In other words, we relied directly on the flattened array of 150,528 pixels (224 pixels × 224 pixels × 3 color channels). This would allow us to understand how deep neural network models helped in processing and extracting relevant features from photographs. For the next five sets of features (see lower half of Figure 2a), we de-identified each photograph by segmenting the face and body (Lugaresi et al., 2019), dilated the mask to obscure edge information, and filled the segmented area using a simple inpainting technique (Bertalmio et al., 2001). The de-identified photographs were entered into the same pipeline as the previous five to extract the final embedding scores (see lower half of Figure 2a).

----- INSERT FIGURE 2 ABOUT HERE -----

We passed every photograph in our sample through the deep neural network models to extract the embedding scores at the end of each model right before the classification block (see Methods). All the models were pre-trained using the VGG-16 deep neural network architecture (see Figure 2b). We deliberately did so to rule out any alternative explanations regarding the design of the networks. The embedding scores were used as inputs to a 10-fold grouped cross-validation pipeline, which fitted a singular-value-decomposition (500 parameters) and logistic regression model to predict the final outcome, similar to previous research (Kosinski, 2021; D. Wang, 2021; Y. Wang & Kosinski, 2018).

Data availability. Data used for this study is available at https://osf.io/4zsn2/.

Code availability. Code used for balancing data, extracting embedding scores, and training models are available at https://osf.io/4zsn2/.

Results

To test the effectiveness of non-biometric, behavioral features extracted from online-posted facial photographs in predicting human traits, we compared the performance of image (ImageNet) (Deng et al., 2009) and scene recognition (Places-365) (Zhou et al., 2018) models to that of a facial recognition model (VGGFace) (Parkhi et al., 2015) in classifying the demographic information in two public datasets (see Methods). We report area under the receiver-operating-characteristic curve (AUROC) in Figure 3 and other common evaluation metrics and confidence intervals estimated by the bootstrap method in Tables 2 and 3. We report the performance of the models in the two datasets separately.

We first assessed the performance of classifications using original photographs (Figure 3a and Table 2a). Overall, both gender and age were classified at AUROC level above random-chance by all models including the baseline (flattened-image). For gender, both the benchmark models, VGGFace (aligned) and VGGFace (unaligned), classified gender close to perfect, replicating results in previous biometric studies (D. Cao et al., 2011; Priadana et al., 2020). In comparison, ImageNet and Places-365 came in very close, followed by the baseline. This pattern is similar for age classification, albeit slightly lower across the board compared to gender classification. That said, age was binned at an arbitrary threshold, thus a more optimized preprocessing pipeline would produce a better result. However, note that achieving state-of-the-art performance is *not* the goal of this study; instead we aim to evaluate the performance of non-biometric models against the benchmark and baseline.

----- INSERT FIGURE 3 ABOUT HERE -----

Surprisingly, even the baseline performed extremely well. This might be due to the very distinctive superficial differences between the prediction classes, making the task easy. Thus, we turned to a conservative test, by which the person in each photograph was completely de-identified. As described in the preprocessing steps (Figure 2a), we ensured that all identifiable information in the photographs was removed. We assessed whether our models could continue to classify demographic traits. The results are reported in Figure 3b and Table 2b.

----- INSERT TABLE 2 ABOUT HERE -----

When photographs were de-identified, demographic traits continued to be classified at levels much higher than random-chance. The benchmark models, which previously achieved remarkable performance, consistently performed the worst among the deep neural network models. The baseline also dropped to almost random-chance level, demonstrating that the tasks here were more difficult than the previous ones. Despite the tasks being difficult, the non-biometric models consistently out-performed the facial recognition models in all classification tasks. The results here demonstrate that when photographs were masked and de-identified, non-biometric deep neural network models would out-perform facial recognition models in the classification of basic human traits.

To verify the robustness of our results, we tested the above analysis using another public dataset, UTK-Face (Z. Zhang et al., 2017), as well as using an additional human trait—race—to demonstrate the generalizability of our effect. Using original photographs, the performance by the benchmark models was unsurprisingly the greatest in all classification tasks (see Figure 3c and Table 3a). This was followed by the non-biometric models. All deep neural network models performed much better than the baseline.

Turning to the more conservative test using de-identified photographs (see Figure 3d and Table 3b), the classification performance for deep neural network models dropped across the board, but was still much higher than the baseline and random-chance. In this case, ImageNet outperformed the benchmark models. Places-365 also achieved high performance, some cases surpassing the benchmark models. Finally, the baseline dropped to almost random-chance, showing that the tasks here were very difficult. These results provide yet more evidence that models not related to biometric-processing, as well as using de-identified photographs, can easily classify human traits from photographs. All in all, we consistently illustrated a serious case of privacy violation not covered by the scope of existing biometric privacy regulations.

----- INSERT TABLE 3 ABOUT HERE -----

Limitations

One may wonder what the mechanisms were that enabled image and scene recognition models to predict human traits. While it is beyond the scope of this study to explore this question, given existing social-psychological research, it is likely that objects in the photographs (Ebert et al., 2021; Torralba et al., 2008; D. Wang, 2021), as well as the context or location where the photographs were taken (Ebert et al., 2021; Matz & Harari, 2020), were robustly related to human traits or behavioral outcomes. These behavioral patterns were encoded in all the pre-trained models, including the facial recognition model, and assisted in predicting the classification outcomes. Future studies could unpack the phenomenon observed here using more human interpretable methods (Rudin, 2019). Future studies could also evaluate the degree to which such behavioral or non-biometric features were present in facial recognition models.

Another limitation is the fact that we only experimented on basic demographic traits and not more personal ones such as sexual (D. Wang, 2021; Y. Wang & Kosinski, 2018) and political orientation (Kosinski, 2021). As mentioned, given our ethical concerns, we believed it unnecessary to obtain more sensitive information to demonstrate this effect. The fact that demographics can be easily detected in people's facial photographs, and without the use of biometric information, demonstrates that people who value privacy in such domains are already at risk of privacy loss. For example, minorities such as Muslims continue to be targets in racial profiling (Crawford et al., 2021). Nonetheless, to ensure the generalizability in other human traits and reproducibility of this research, we welcome future researchers, who similarly wish to raise awareness of privacy concerns, to replicate our findings using our open-source codes (see Code Availability).

General Discussion

In this research, we demonstrate an alarming case, in which one of the strictest privacy regulations concerning facial recognition—BIPA—would fail at protecting people's privacy. Specifically, we show how image and scene recognition models, which are pre-trained in tasks far from biometric processing, can predict human traits from online posted facial photographs at levels almost comparable to facial recognition, bypassing the jurisdictions of current biometric privacy regulations. Furthermore, we demonstrate a conservative test by which this phenomenon was still present when the face and body in each photograph were masked and de-identified. In such case, non-biometric models virtually out-performed facial recognition models in classifying these important traits. Overall, the results in this research illustrate how biometric-equivalent information was almost not necessary in exposing human traits, rendering biometric protection regulations ineffective at protecting people's privacy in online posted photographs.

Recently, big technology companies, such as IBM, Amazon, Microsoft, and Facebook, consecutively halted the use of facial recognition technology (Palmer, 2021), yet according to the

findings in this research, the danger of privacy loss still looms large. Facebook announced that the company has shut down its facial recognition system as well as deleted all facial identifiers of its users (Hill & Mac, 2021). But since then, there has not been any public announcements about how the company would handle raw photographs. As most technology companies' business models depend on the use of consumers' data (Hartmann et al., 2016; West, 2019), if companies can still extract useful information from these data without the use of facial recognition and without reliance on biometric-equivalent information, the danger of privacy loss would still be present in our society.

We hope the current research contributes to the urgent ongoing discussion on how privacy regulations should be formulated in places where no privacy regulation exists, or reformed in places where only biometric privacy acts are in place. By unpacking the process and drawing some theoretical understanding on how human traits in facial photographs were revealed, we aim to shine a light on the current situation of privacy in the face of deep learning. Where administrators' resources are sufficient, we recommend that policymakers should consider regulating both the input (facial photographs) and output (biometric information) of facial recognition. We also recommend that policymakers consider incorporating behavioral data into the scope of privacy laws because these data might reveal certain sensitive information more effectively than biometric data. Lastly, we hope policymakers can consider asking companies to publicly disclose their usage of facial photographs as well as other digital data to relevant auditing bodies with expert knowledge of privacy protection.

Chapter 3

The Past is Better than the Future of Organizations Research

Conceptualizing the Theory of Artificial Intelligence Application as a Decision-Making Process

Introduction

From smart phones to household appliances, artificial intelligence increasingly impacts our lives in ways once thought unimaginable. Utilizing decision-making systems based on computer algorithms, more organizations are incorporating automation technologies, replacing traditional systems based on individual or organizational intelligence. This trend, catalyzed by the recent Covid-19 pandemic, saw a growing number of companies adopt cloud-based, deep learning tools (Obrad & Circa, 2021); some employed automated interview systems that computerize hiring decisions, while others rolled out productivity monitoring algorithms that manage employees working from home. If these developments herald the new digital revolution changing how we live, work, and play, we, as organizational scholars, must ask: What will be the future of organizations?

As the study of the future of organizations becomes a hot topic in organization and management research (Bailey et al., 2022), three perspectives have emerged, namely the automation, augmentation, and automation-augmentation paradox perspective (Raisch & Krakowski, 2021). Table 1 summarizes these perspectives. Since their emergence, advocates of the three camps have been engaged in heated conversations with the other camps over which perspective would most accurately describe the future of organizations (Lindebaum & Ashraf, 2021), and which path is normatively more beneficial to organizations, societies, and what may come beyond (Raisch & Krakowski, 2021).

However, these discussions often have ended unproductive, because the current scholarship seems to be fragmented at best and subjective at worst (for an example, see conversations between

Lindebaum & Ashraf, 2021 and Leavitt et al., 2020, 2021). It would not be an exaggeration to equate the impressions that many researchers portray, to scenes in a science-fiction novel, leaving one questioning whether such impressions are realistic and reflective of the actual phenomenon. Having many redundancies in theorization is not detrimental, but if noise turns away future scholars, or dampen optimism in current scholars, it would be a zero-sum game for everyone in this field (Colquitt & Zapata-Phelan, 2007).

We argue that the current fragmentation in artificial intelligence application theory would be improved by removing a few roadblocks. First, camps have not been communicating using a similar set of language, and therefore would greatly benefit with a bridging theme and a theory that unifies disparate constructs. For example, contributors to the automation perspective typically draw from technical research such as machine learning, computer science, and information systems research (Frey & Osborne, 2017). On the other hand, the augmentation perspective is situated mostly at the strategy (Choudhury et al., 2019, 2021), psychological (Logg et al., 2019), and decision-making (De Cremer & De Schutter, 2021; Lebovitz et al., 2022) literature, and a select few studies have convered the organizational perspective (Balasubramanian et al., 2020; Puranam, 2021).

Second, camps use various units of analysis, similar to how management scholarship is broken into micro and macro positions (Rousseau, 2011). For example, the automation perspective is mostly interested in micro-level behaviors of individuals and machines, and how they linearly aggregate to organizational or societal outcomes, ignoring meso- and macro-level mechanisms of organizing (Felten et al., 2021; Frank et al., 2018; Frey & Osborne, 2017). While the automation-augmentation paradox perspective is concerned with the meso- or macro-level mechanisms, there is less discussion of how micro-level, cognitive mechanisms interact (Raisch & Krakowski, 2021). Lastly, despite all perspectives having some interest in intelligence and how intelligent systems contribute to organizational outcomes, they neglect a rich tradition of research in organizations and organizational intelligence (Levitt & March,

1988; March & Simon, 1958; Walsh & Ungson, 1991). All these limitations result in a set of mid-level theories that would fail to generalize beyond time and space, and interest audience beyond the field.

In this work, we argue that the past is in fact better than the future in the theory of artificial intelligence application in organizations. This is because pioneering research in both domains has much to offer. By reviving the pioneering research, we contribute to the discourse of artificial intelligence application twofold: (1) conceptualizing tasks as decision-making, forging a level playing field for all perspectives and positioning the unit of analysis as a variable not limited to any levels, and (2) re-introducing artificial intelligence, and introducing organizational intelligence as forms of decision-making systems that prioritize the "human" element over substantive or formal rationality.

We began by reviewing the three major perspectives in artificial intelligence application. To conceptualize tasks as decision-making, we draw upon the bounded-rationality framework proposed in Simon (1947). We review how pioneering research in individuals, organizations and artificial intelligence all emerged, coordinately, in relation to bounded rationality in decision-making. To provide a working theoretical lens on how we can view issues in artificial intelligence application, we invoke the means-ends chain proposed in March and Simon (1958). We provide some non-exhaustive spiel of examples how this lens can be adopted in our discourse and how it connects with many influential ideas in our domain. We conclude by pointing out how the use of our framework contributes to artificial intelligence application as well as research in management and organization.

Main

Perspectives of Artificial Intelligence Applications in Management

Three perspectives—automation, augmentation, and automation-augmentation paradox—have recently emerged in management and organization research; they are outlined in Table 1. In general, despite many disagreements, all three perspectives would agree on the fact that humans and machines are capable of accomplishing organizational tasks. However, they disagree in their assumptions of how

humans and machines could accomplish these tasks, as well as how these differences would project the role machines play in organizations. In this section, we briefly review these perspectives.

----- INSERT TABLE 1 ABOUT HERE -----

The Automation Perspective. This perspective, also referred to as the human "out of the loop" approach, predicts a future when certain tasks and occupations, otherwise only accomplishable through humans, will be replaced by machines (Brynjolfsson & McAfee, 2014). For example, Frey and Osborne (2017) identified three sets of bottlenecks to computerization—perception and manipulation, creative intelligence, and social intelligence. Using these bottlenecks, the authors estimated the probability that 702 occupations would be substituted by automation, and found that transportation, administration, and repetitive jobs are more likely to be at risk.

Since Frey and Osborne's (2017) paper has been published, a train of research is published using similar framework and methodology. For example, Felten et al. (2021) created an Artificial Intelligence Occupational Exposure (AIOE) measure, which evaluated each occupation's susceptibility to automation using a set of 10 artificial intelligence applications. They aggregated the score from the occupational level to the industry level to predict the trajectory of industries in the wake of artificial intelligence. Similarly, Frank et al. (2018) extended Frey and Osborne's (2017) computerization predictions to geographical locations in the United States, and found that small cities are more at risk by automation.

Most of the research in this perspective is based on the assumption that machines are either capable of or limited in completing certain tasks compared to humans, but most capabilities and limitations are defined rather subjectively. For example, Frey and Osborne (2017) wrote: "we subjectively hand-labelled 70 occupations, assigning 1 if automatable, and 0 if not." Since the publication of Frey and Osborne (2017), research has questioned their predictions. Autor's (2015) title,

"Why Are There Still So Many Jobs," aptly summarizes this phenomenon. In a large-scale, near universal analysis of online job vacancies, Acemoglu and Restrepo (2020) found no detectable automation effect on the aggregate labor market.

The Augmentation Perspective. An augmentation perspective, or the human "in the loop" approach, refers to the scenario where, instead of replacing humans, machines would assist humans in performing tasks (Brynjolfsson & McAfee, 2014; Davenport & College, 2016; Wilson & Daugherty, 2018). Augmentation is defined as "a process of enlargement or making something grander or more superior" (Lebovitz et al., 2022, p. 127). By expanding each other's knowledge, expertise, and capabilities, human-machine augmentations are predicted to help organizations reap superior performance (Lebovitz et al., 2022; Wilson & Daugherty 2018).

Typically, this approach is based on the assumption that machines are still vastly bounded in their capabilities, thereby warranting human expertise in their applications (Stohl, 2016). For example, Lebovitz et al. (2022) found that whether or not a diagnostician integrated claims made by artificial intelligence in judging diseases is moderated by the uncertainty arising from opacity in the machine's decision-making process and the ability for the diagnostician to enact interrogation practices. This finding represents a great majority of writings in augmentation perspective, which are concerned with the configuration of human-machine collaboration (Balasubramanian et al., 2020). Some possible dependent variables include whether organizations adopt human-machine collaborations parallelly, sequentially, or at all (Puranam, 2021).

One limitation in this line of work is that despite artificial intelligence application being conceptualized as an important variable affecting organizational intelligence, discussion of the relationship of artificial intelligence to organizational intelligence, as well as mechanisms connecting artificial to organizational intelligence and organizational learning, are lacking. For example,

Balasubramanian et al. (2020) calls for a "deeper conversation about the risks and benefits of ML [machine learning], and the roles of humans therein."

The Automation-Augmentation Paradox Perspective. This perspective focuses on the temporal and spatial process of artificial intelligence applications (Raisch & Krakowski, 2021). It proposes that automation and augmentation cannot be neatly separated from each other; instead they are dynamically interdependent, and the use of any one-sided strategy is not sufficient to ensure positive organizational and societal outcomes. For example, humans would first augment the development of AI capabilities because humans can use their expertise to "evaluate, select, and complement machine outputs" (Raisch & Krakowski, 2021, p. 196). Sequentially, such capability would lead to the automation of the task on a temporal scale. The automation of one task may spill over to other tasks, leading to an increase in automation in the spatial scale. As this virtuous cycle continues, augmentation and automation would interdependently improve organizational and societal outcomes. On the other hand, choosing one strategy over another may lead to a vicious cycle, due to the adoption of a partial solution.

This perspective is valuable in shifting our focus from a static view to a temporal and spatial process of artificial intelligence application (Raisch & Krakowski, 2021). Indeed, technology advances quickly, very likely resulting in research failing to generalize in a few years. For example, Bailey et al. (2022, p. 1) recommended that scholars "treat these new technologies as 'emerging' because their uses and effects are still varied and have yet to stabilize around a recognizable set of patterns and because the technologies themselves are, by design, always changing and adapting." We aim to build on this "process-focused" (Raisch & Krakowski, 2021) and "relational" view of artificial intelligence application (Bailey et al., 2022). In addition, we aim to argue that one variable could remain constant in organizations—decision-making—and would unify the process-relational view of artificial intelligence application. In the next section, we introduce decision-making as the unit of analysis, and call for a unified theory that is built around this tradition.

Decision-Making As The Unit of Analysis

We call for a more focused approach that considers decision-making as the unit of analysis (Cyert & March, 1963). As Simon (1947, pp. xiii–xiv) posited, "[d]ecision making is the heart of administration,... [t]he vocabulary of administrative theory must be derived from the logic and psychology of human choice." Since the publication of Simon's theory of management, few other works have matched its extraordinary influence, with its reach expanding well beyond its original domain (Gavetti et al., 2007, 2012). We aim to connect with this rich tradition of research by focusing on decision-making, and to position the theory of artificial intelligence application at the heart of management and organizational research (Bailey et al., 2022). We offer a few reasons in this section.

First, theorizing around the decision-making tradition allows us to tap into a readily available and well-developed set of theoretical tools (Cyert & March, 1963; Gavetti et al., 2007; Gavetti et al., 2012; Luan et al., 2019). Concentrating on decision-making, the behavioral theory of the firm offers a set of open and versatile theoretical apparatus that "proved to be a source of strength, allowing a broad community of scholars to build on these ideas and offering opportunities for further enrichment of these ideas" (Gavetti et al., 2012, p. 29). Therefore, focusing on decision-making allows us to tap into many practical benefits in theory building. Conversely, adopting the theorical apparatus to a novel phenomenon would also contribute significantly to the relevance and continued interest of this classical theory in the current digital age.

Second, focusing on decision tasks also revives the pioneering research in artificial and organizational intelligence. In their introductory paragraphs, Raisch and Krakowski (2021) rightfully cited works by Simon and colleagues (e.g., Newell et al., 1958; Simon & Newell, 1961). It is not a coincidence that Simon laid the foundation for both the classical theories in organizations and artificial intelligence. As we will highlight in our subsequent discussion, pioneering artificial intelligence research, which had

remained dormant since the 1960s (Buchanan, 2005), as well as organizational intelligence research has much to offer.

Third, most theories in artificial intelligence application is partly built upon the premise that humans and machines commit bias in the application of artificial intelligence. However, despite bias being a crucial building block in these theories, the definition of bias is not clear. In fact, the use of "bias" is misleading because in decision-making under uncertainty (i.e., prediction), the bias-variance framework argues that bias is only one of the three components of any prediction error, the other two being variance and randomness (Luan et al., 2019). We believe this vagueness arose essentially from the lack of boundary in what tasks are. If we limit tasks to decision-making, outcomes of such tasks could be readily evaluated using measurements that already exist in such tradition.

Bounded Rationality in Human Intelligence

In his seminal work, "Administrative Behavior," Simon (1947) proposed that humans are boundedly rational. According to Simon (1969), "What a person cannot do he or she will not do, no matter how strong the urge to do it…" (p. 28). Grounded in realism and drawing from psychology, Simon (1969) proposed that decision-makers, despite being self-interested, are vastly bounded through knowledge and computation. A real decision-maker cannot achieve rationality simply because they do not have the ability to do so. This proposal was a fervent response to the then dominant rational-agent model of decision-making in economics, which described human agents as "omniscient demons" (Simon, 1976), who are "motivated by self-interest and completely informed about all available alternatives" (Scott & Davis, 2006).

Given bounded rationality, a limitedly rational decision-maker would engage in a decision-making process called satisficing. The word "satisficing" is formed from the words "satisfy" and "suffice" (Simon, 1947). A typical individual would search for alternatives to a decision, but stop right after the first satisfactory alternative. For example, in personnel selection, a satisficing recruiter would start to

interview candidates even before the application deadline ends. Even if the recruiter waited for the deadline to end, there might be cases when suitable candidates did not apply for the job because the job advertisement failed to reach them. In such cases, a satisficing recruiter typically would not actively search for those candidates and would start to hire the first satisfactory candidate. Lastly, instead of relying on all information for each candidate, a satisficing recruiter would rely on stages of evaluation, screening candidates by batches and in rounds. Overall, "Individuals are boundedly rational because they know but a tiny fraction of the possible choice alternatives and their values" (Gavetti et al., 2012, p. 5).

After discovering bounded rationality as a more realistic model of decision-making (Simon, 1947), Simon shifted his focus to other forms of intelligence, contemplating the question: how do other forms of intelligence, in relation to human intelligence, address issues in rationality in decision-making? Subsequently, most of his research can be roughly categorized into two domains: artificial intelligence with colleagues, eminently Allen Newell (see Newell et al., 1959), and organizations or organizational intelligence with colleagues, notably James March (see March & Simon, 1958). In the next sections, we review these works and draw connections to our current discussion of artificial intelligence applications.

Pioneering Research in Artificial Intelligence

Gugerty (2006) argued that Newell et al.'s (1959) artificial intelligence algorithm "was perhaps the first working program that simulated some aspects of peoples' ability to solve complex problems." In 1956, Simon persuaded his university (then known as the Carnegie Institute of Technology) to purchase its first computer, the IBM 650. It was housed in the basement of the business school. There, Simon and Newell created the General Problem Solver, which became one of the earliest artificial intelligence programs. The program was in fact created to solve a decision-making task. The Rand Corporation adopted it in its missile defense system, in which the computer program provides responsive and timely decisions during a missile strike.

Unfortunately, as Raisch and Krakowski (2021) explained, early artificial intelligence research soon entered into a state of dormancy, widely known as the Artificial Intelligence Winter As we enter into a season of revival in artificial Intelligence research, it is timely to look back and understand what led to the winter of artificial intelligence research. We believe one possible reason is a myopic interpretation of the intentions of Simon's pioneering research in artificial intelligence.

Because Newell and Simon's research in artificial intelligence was conceptualized as a decision-making system, on surface, the research seems to be concerned with problems of bounded rationality. Indeed, machines improve computational efficiencies, response time to decisions, and ability to conduct search for alternatives and information, compared to human decision-makers. However, these improvements were in fact not the entire purpose. "[T]he computer, as a piece of hardware, or even as a piece of programmed software, has nothing to do directly with the matter" (Simon, 1969, p. 126). We will highlight in our subsequent sections how bounded rationality ultimately gave way to procedural rationality in organizational intelligence research. While people discredited early artificial intelligence programs for falling short of expectations and relying on rule-based (instead of learning-based) algorithms, these programs were in fact, not *entirely* created to demonstrate accuracies or efficiencies in computations. What, then, are the other purposes?

According to Simon, if we considered all man-made systems as artificial, the study of artificial intelligence would become a study of design[11]—the reasons behind design choices (procedural rationality) and how designs adapt to external environments (ecological rationality) instead of the design outcomes (substantive rationality). Sadly, this misinterpretation, or the focus on outcomes, is common. In the contemporary academic community, for example, a psychologist might use machine

[11] The other overlooked purpose in Newell et al.'s (1959) pioneering artificial intelligence research was to model human decision-making and to study the human mind. Despite Simon knowing all too well the bounds and limits of human rationality, the early artificial intelligence systems in Newell et al. (1959) were, as a matter of fact, based on the human thought process.

learning to study human cognition by comparing accuracies in two study conditions, but their intention would easily be discredited due to the fact that their accuracies in predictions were too low. Simon (1969, p. 126) wrote: “Consequently we as designers, or as designers of design processes, have had to be explicit as never before about what is involved in creating a design and what takes place while the creation is going on.”

Overall, we identify three valuable lessons from pioneering research in artificial intelligence. First, its conceptualization of artificial intelligence as decision agents helps us to connect artificial intelligence research to the original problem of bounded rationality as well as relationships to other forms of intelligent systems such as organizational intelligence. As explained in Simon (1991), “Artificial intelligence was born in close connection with management science, grew apart from it, and is now forming new links with it, as well as with the other disciplines that have come together in cognitive science.” Comparing capabilities and limitations of the different intelligence systems is one goal; other goals include modeling systems based on their counterparts (e.g., see Csaszar & Steinberger, 2022).

Second, the conceptualization of artificial intelligence research as a study of design and design process shifted the focus from the outcome of design to the mechanisms and process in design. It also gave rise to a temporal, iterative view of system design, which is crucial for our current understanding of the application of artificial intelligence. As Bailey et al. (2022) rightfully pointed out, technology is constantly changing. If we view artificial intelligence as playing a crucial relational role in a bigger system of organized intelligence, we will achieve greater openness to possibilities, descriptive power, and generalizability of our theories.

Third and relatedly, anchoring too deep on the accuracies, efficiencies, and rationalities of artificial intelligence makes us lose sight of the joy and human elements of research. This warning is acknowledged in Lindebaum et al. (2020), who argued that “decision making premised on formal rationality is not based on the qualities of the individual concerned—neither the judgment of the

decision maker nor the specific conditions of the decision maker—but, rather, is predicated on universalism and calculation with reference to formal rules and regulations."

Organizational Intelligence and Procedural Rationality

March & Simon (1958) proposed that organizations are an alternative form of intelligence in relation to bounded rationality in individual decision-making. An organization is a distributed system of human decision-makers, organized to capitalize on individual intelligence such that "[a]n organization may be smarter than its individual members" (Glynn, 1996, p. 1091). According to Simon (1969, p. 42): "business organizations, like markets, are vast distributed computers whose decision processes are substantially decentralized. The top level of a large corporation, which is typically subdivided into specialized product groups, will perform only a few functions."

Organizations engage in a form of distributed intelligence or cognition; other forms of organizational intelligence include cross-level and aggregated intelligence (for a review, see Glynn, 1996). Distributed organizational intelligence is "embedded in the organization's systems, routines, standard operating procedures, symbols, culture, and language" (Glynn, 1996, p. 1091). These systems "both simplify decisions and support participants in the decision they need to make" (Scott & Davis, 2006 p. 53).

Simplification is achieved through limiting the scope of the decision. If the decision process is viewed as means (method in achieving expectations from a decision) and ends (expectations of a decision), simplification parcels out the ends. For example, organizations distribute decision-making responsibilities into specialized, hierarchical divisions, limiting the space of information search, simplifying the decision topic-wise. For example, organizations are divided into marketing, finance, accounting, human resources, and so on. Each member in the sub-unit requires only knowledge in their own domain. Organizations also decentralize decision-making hierarchically. Hierarchical structure helps to break tasks such that "[those] closer to the top make decisions about what the organization is going

to do; those in lower positions are more likely to be allowed to make choices as to how the organization can best carry out its tasks" (Scott & Davis, 2006, pp. 53–54).

Support is given by formalized structure. Besides providing them with resources, tools, and the necessary information to make appropriate decisions, organizations support their participants by enacting formal rules, guidelines, standard operating procedures, and routines. For example, in novel and uncertain situations, the alterative space is infinite, and thus existing organizational knowledge that exists in the organizational memory would support each individual decision-maker.

As the focus is shifted from decisions per se to how decisions are deliberated, the act of organizing is consequently shifted from substantive rationality, which is rationality in the outcome of the decision, to procedural rationality, which is rationality in the process of making decisions (Simon, 1976). By paying more attention to the process, we can hope for more rational outcomes. Note that procedural rationality in decision-making exists not only at the organizational level but at the individual level, making it a variable that is not limited to any level of analysis. If the reader is convinced that organizations are intelligence systems used to address decision-making issues, the next question to ask is: how should we study it? In the next section, we propose a working framework and a bag of non-exhaustive theoretical tools to study organizational intelligence in relation with artificial intelligence.

Means-Ends Chains of Decision-Making

Unbeknownst to many, the means-ends analysis is a common theme in Simon's study of individuals, organizations, and artificial intelligence. Consider the following quote from Simon (1976, p. 68), which draws similarity between human and computer intelligence:

> Like a modern digital computer's, Man's equipment for thinking is basically serial in organization. That is to say, one step in thought follows another, and solving a problem requires the execution of a large number of steps in sequence. The speed of his

> elementary processes, especially arithmetic processes, is much slower, of course, than those of a computer, but there is much reason to think that the basic repertoire of processes in the two systems is quite similar. Man and computer can both recognize symbols (patterns), store symbols, copy symbols, compare symbols for identity, and output symbols. These processes seem to be the fundamental components of thinking as they are of computation.

Here, Simon described a process largely similar to the modern day decision-tree analysis in machine learning. Decision process is decentralized into branches and nodes. Each branch represents a possible reality and each node a decision, whose outcome would lead to a subsequent decision. To study this in organizations, March and Simon (1958) proposed using goals of organizations "as the starting point for the construction of means-ends chains" (Scott & Davis, 2007, p. 54). This is also similar to how most computer scripts are executed, from the top to the bottom.

For example, how does an entrepreneur achieve profit-maximization after deciding to go into, say, the personnel selection field? Following March and Simon (1958), when evaluating artificial intelligence application in organizations, scholars can choose to take organizational goals as given[12] (i.e., deciding to go into personnel selection) and examine subsequent blocks in the means-ends chain of decision analysis (i.e., how to execute this goal). Once a goal is chosen, each level of the means-ends chain can be analyzed by looking at the input and output of each level. The end is the mean of the previous level, the mean is the end of the next level, and so on.

The use of means-ends chains of decision is practically beneficial in the analysis of artificial intelligence application because we can leverage on much existing theory that connects the "ends"

---

[12] To the best of our knowledge, artificial intelligence has not yet been employed to decide organizational goals.

component of decisions to organizational outcomes. For example, the value-factual dichotomy differentiates the ends and relates them to organizational hierarchy. Specifically, value premises of goals are the desires, hopes, and aspirations arising out of a decision, and factual premises are representative observations of the world (March & Simon, 1958). This distinction is relevant to our discussion, because while most perspectives draw such differentiation, its implication for organizational design is not clearly stated. One design consideration is that decision-makers closer to the top of the organization typically make more value judgments, and those closer to the bottom make more factual judgments (Scott & Davis, 2006. If computerized decision-making is mostly factual, because machines are capable of consolidating and discovering patterns in big data (Choudhury et al., 2021), the decision of whether to support one political wing over another will still be formed through "the [human] members of the organizational coalition" (Cyert & March, 2002, p. 164).

Apart from connecting artificial intelligence application research to organizational design, the means-ends chain analysis also allows organizational scholars to tap into a rich bag of apparatus that looks at the "means" component. A few non-exhaustive examples include theories on expectations (Cyert & March, 1958), experiential search (Gavetti & Levinthal, 2000), organizational learning (Levitt & March, 1988), and exploration-exploitation framework (March, 1991). In simple terms, after a goal is chosen, the decision-maker needs to make decisions on how to accomplish the goal. To decide, the entrepreneur would have to engage in a process of forming expectations or the prediction of future events, assuming no prior knowledge (Cyert et al., 1958; Cyert & March, 1963, 2002). In economics, expectations are given—no resources are needed to predict the outcome of a decision (Simon, 1969; Cyert & March, 1963). In human decision-making, "expectations of the attainable define an aspiration level that is compared with the current level of achievement" (Simon, 1969, p. 30). They are hopes and wishes of the individual. In organizational decision-making, expectations are formed also through the preferences of the sub-unit at the level of analysis.

As expectations affect choice, which in turn affects the outcome of the decision, humans (and organizations) engage in what Simon (1969) called feed forward mechanism. If achievements fall short of expectations, the decision-maker would experience negative feedback, and the decision-maker would continue to search for more satisfactory outcomes. Otherwise search is stopped. In organizations, however, positive feedback is encoded into routines in a process called organizational learning (Levitt & March, 1988). Thus, individuals and organizations adapt to the environment by progressively acquiring more knowledge, stored either in individual or organizational memory that could be retrieved to guide future behaviors (Levitt & March, 1988).

Organizational learning represents a temporal and spatial process that is vastly similar to the process described in the paradox perspective of artificial intelligence application (Raisch & Krakowski, 2021). For example, the organizational learning literature would predict organizational failure in the long-run if artificial intelligence is adopted using the automation strategy. This is because routines are detrimental to organizations due to the lack of exploration of new possibilities (March, 1991). Routines are also harmful to a firm's adaptation to changing environments. Gavetti and Levinthal (2000, p. 113) argued that, "[c]hanging a cognitive representation itself can act as an important mode of adaptation, effectively resulting in the sequential allocation of attention to different facets of the environment." Following this logic, increasing automation would lead to a tension requiring a change in cognitive representation, that is, more augmentation in the automation-augmentation cycle.

Overall, in this section, we recommend adopting the means-ends chains of decision-making when theorizing the application of artificial intelligence for several reasons. This approach separates decision-making into inputs and outputs. We began by arguing that organizational goals can be taken as the starting point in this analysis. We then demonstrated that ends can be separated into useful dimensions such as the value-factual premise. Means can be further studied using theories such as

expectations, representations, and information search. Lastly, the temporal and spatial aspect of means-ends chains can be easily connected to burgeoning theories in artificial intelligence application.

Conclusions

In this article, we began with three bottlenecks in existing perspectives on the application of artificial intelligence in organizations. These bottlenecks are a lack of similar language, inconsistent units of analysis, and a disconnect with organizational intelligence research. Motivated by these limitations, we propose a unified theory and analysis centered around decision-making and the bounded rationality framework. We argue that this framework would be a level playing ground for all perspectives because it creates a clear dependent variable—decision-making—as well as a set of existing theoretical apparatus to draw from. We then reviewed human, artificial, and organizational intelligence research and provided a working framework employing means-ends analysis. We believe this unified lens of looking at the application of artificial intelligence would lead to more fruitful and productive discourse between the various perspectives. This framework would also revive pioneering research in both artificial intelligence and management research. In closing, we believe that the past would tremendously shape our future for organizations and organizational research.

Tables

Introduction

Table 1

A Summary of Artificial Intelligence Application in Social Psychological Research

Input	AI Application	Output	Citation
Facial images	DNN / PCA	Criminal behavior	Wu & Zhang, 2016
Facial images	PCA	Leadership emergence	Stoker et al., 2016
3D facial images	PCA	Personality	Hu et al., 2017
Facial images	API	Behavioral tendencies	Kosinski, 2017
Facial images	DNN	Sexual orientation	Y. Wang & Kosinski, 2018
Facial images	API	Behavioral tendencies	D. Wang et al., 2019
Facial images	DNN	Personality	Kachur et al., 2020
Survey Responses	DNN	Unethical behavior attitudes	Sheetal et al., 2020
Facial images	DNN	Political orientation	Kosinski, 2021
Survey Responses	DNN	Cultural change markers	Sheetal & Savani, 2021
Facial images	DNN	Sexual orientation	D. Wang, 2022
Facial images	DNN	Demographics	D. Wang, Under review

Note: AI stands for artificial intelligence, DNN stands for deep neural networks, PCA stands for principal component analysis, API stands for application programming interface

Chapter 1

Table 1

Sample breakdown by users and number of facial images for studies 1a, 1b, 1c, 2a and 2b

	Women		Men	
	Lesbian	Heterosexual	Gay	Heterosexual
Unique users	5170	5170	2562	2562
Total images	10800	10800	5081	5081
Mean age (SD)	27.8 (4.5)	27.8 (4.5)	25.8 (4.3)	25.8 (4.3)
Users with:				
1 image only	2259	2259	1189	1189
2 images only	1266	1266	641	641
3 images only	972	972	454	454
4 images only	400	400	177	177
≥5 images	273	273	101	101

Table 2

Comparison of the means of self-presentational facial attributes by sexual orientation and gender for study 1a

	Women (N = 10,340)										
	Heterosexual			Lesbian			Significance Test				
	Mean	95% CI		Mean	95% CI		*t*	95% CI		*p*	*d*
Fac. exp.:											
Neutral	0.254	[0.24,	0.26]	0.325	[0.32,	0.34]	10.408	[0.06,	0.08]	<.001	0.205
Happiness	0.622	[0.61,	0.63]	0.522	[0.51,	0.53]	-12.804	[-0.12,	-0.08]	<.001	0.252
Anger	0.015	[0.01,	0.02]	0.016	[0.01,	0.02]	1.263	[-0.00,	0.00]	.206	0.025
Disgust	0.016	[0.01,	0.02]	0.020	[0.02,	0.02]	1.938	[-0.00,	0.01]	.053	0.038
Surprise	0.049	[0.05,	0.05]	0.060	[0.06,	0.06]	3.457	[0.00,	0.02]	<.001	0.068
Sadness	0.031	[0.03,	0.03]	0.039	[0.04,	0.04]	3.251	[0.00,	0.01]	<.001	0.064
Head pose:											
Roll (abs)	0.031	[0.03,	0.03]	0.030	[0.03,	0.03]	-2.300	[-0.00,	-0.00]	.021	0.045
Yaw (abs)	0.458	[0.45,	0.46]	0.447	[0.44,	0.45]	-2.400	[-0.02,	-0.00]	.016	0.047
Pitch (abs)	0.551	[0.54,	0.56]	0.518	[0.51,	0.52]	-7.143	[-0.04,	-0.02]	<.001	0.140
Smiling	0.680	[0.67,	0.69]	0.576	[0.57,	0.59]	-13.441	[-0.12,	-0.09]	<.001	0.264
Eyes	0.117	[0.11,	0.12]	0.158	[0.15,	0.17]	7.296	[0.03,	0.05]	<.001	0.144
Glasses	0.183	[0.17,	0.19]	0.264	[0.25,	0.27]	11.150	[0.07,	0.10]	<.001	0.219
	Men (N = 5,124)										
	Heterosexual			Gay			Significance Test				
	Mean	95% CI		Mean	95% CI		*t*	95% CI		*p*	*d*
Fac. exp.:											
Neutral	0.394	[0.38,	0.41]	0.441	[0.43,	0.46]	4.392	[0.03,	0.07]	<.001	0.123
Happiness	0.479	[0.46,	0.49]	0.424	[0.41,	0.44]	-4.845	[-0.08,	-0.03]	<.001	0.135
Anger	0.013	[0.01,	0.02]	0.017	[0.01,	0.02]	1.502	[-0.00,	0.01]	.133	0.042
Disgust	0.030	[0.03,	0.03]	0.029	[0.03,	0.03]	-0.308	[-0.01,	0.00]	.758	0.009
Surprise	0.030	[0.03,	0.03]	0.034	[0.03,	0.04]	1.455	[-0.00,	0.01]	.146	0.041
Sadness	0.036	[0.03,	0.04]	0.039	[0.03,	0.04]	0.640	[-0.00,	0.01]	.522	0.018
Head pose:											
Roll (abs)	0.040	[0.04,	0.04]	0.041	[0.04,	0.04]	1.740	[-0.00,	0.00]	.082	0.049
Yaw (abs)	0.392	[0.38,	0.40]	0.408	[0.40,	0.42]	2.400	[0.00,	0.03]	.016	0.067
Pitch (abs)	0.438	[0.43,	0.45]	0.452	[0.44,	0.46]	2.023	[0.00,	0.03]	.043	0.057
Smiling	0.556	[0.54,	0.57]	0.493	[0.48,	0.51]	-5.584	[-0.08,	-0.04]	<.001	0.156
Eyes	0.227	[0.21,	0.24]	0.211	[0.20,	0.22]	-1.655	[-0.04,	0.00]	.098	0.046
Glasses	0.235	[0.22,	0.25]	0.271	[0.26,	0.29]	3.341	[0.02,	0.06]	<.001	0.093

Table 3

Average accuracy afforded by self-presentational facial attributes by gender for study 1b

	Women				Men			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Neutral	54.93%	56.20%	44.72%	49.81%	52.62%	52.73%	50.59%	51.63%
Happiness	55.95%	56.31%	53.08%	54.65%	52.97%	52.89%	54.37%	53.62%
Anger	50.48%	52.39%	10.62%	17.66%	50.35%	51.50%	12.06%	19.54%
Disgust	50.27%	51.31%	10.58%	17.54%	50.02%	50.01%	85.21%	63.03%
Surprise	51.39%	54.13%	18.26%	27.31%	50.10%	50.38%	12.80%	20.42%
Sadness	51.20%	54.18%	15.53%	24.14%	49.75%	49.31%	18.03%	26.41%
Combined	56.21%	56.85%	51.55%	54.07%	52.48%	52.47%	52.58%	52.52%
Roll (abs)	50.44%	50.37%	60.81%	55.10%	51.19%	51.47%	41.80%	46.13%
Yaw (abs)	50.70%	50.68%	51.90%	51.28%	51.17%	51.25%	48.01%	49.58%
Pitch (abs)	53.32%	53.45%	51.37%	52.39%	51.07%	51.11%	49.26%	50.17%
Combined	53.41%	53.55%	51.55%	52.53%	52.17%	52.24%	50.59%	51.40%
Eyes	52.92%	56.32%	26.02%	35.59%	51.13%	50.84%	68.46%	58.35%
Glasses	54.67%	57.70%	35.01%	43.58%	51.81%	52.76%	34.74%	41.89%
Smiling	56.19%	57.17%	49.38%	52.99%	53.22%	53.17%	54.02%	53.59%
Combined	57.85%	57.87%	57.76%	57.81%	53.81%	53.67%	55.66%	54.65%
Combined	58.34%	58.54%	57.12%	57.82%	53.47%	53.48%	53.40%	53.44%

Table 4

Baseline accuracy results by number of images and gender for study 1c

	Women				Men			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Average	64.83%	64.72%	65.18%	64.95%	61.10%	61.21%	60.66%	60.93%
1 Image	63.64%	63.39%	64.55%	63.96%	59.37%	59.44%	58.98%	59.21%
2 Images	65.22%	65.23%	65.17%	65.20%	62.56%	62.31%	63.58%	62.94%
3 Images	65.62%	65.83%	64.98%	65.40%	64.75%	64.03%	67.35%	65.65%
4 Images	65.68%	65.40%	66.57%	65.98%	65.29%	64.91%	66.55%	65.72%
≥5 Images	66.85%	66.91%	66.67%	66.79%	72.28%	70.64%	76.24%	73.33%

Chapter 2

Table 1

Final sample size and by labels and prediction classes

	Gender		Age		Race	
	Label = 0	Label = 1	Label = 0	Label = 1	Label = 0	Label = 1
CelebA-HQ						
By image:	4,009 (50%)	4,009 (50%)	4,009 (50%)	4,009 (50%)	-	-
By person:	1,173 (50%)	1,173 (50%)	1,173 (50%)	1,173 (50%)	-	-
UTK-Face						
By image:	10,221 (50%)	10,221 (50%)	10,221 (50%)	10,221 (50%)	10,221 (50%)	10,221 (50%)
By person:	10,221 (50%)	10,221 (50%)	10,221 (50%)	10,221 (50%)	10,221 (50%)	10,221 (50%)

Table 2

Classification performance using CelebA-HQ dataset

a. Original image						
		VGGFace (aligned)	VGGFace (unaligned)	ImageNet (unaligned)	Places-365 (unaligned)	Flattened-image (unaligned)
Gender	AUROC	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	0.99 (0.98, 0.99)	0.99 (0.98, 0.99)	0.98 (0.98, 0.99)
	Accuracy	0.98 (0.97, 0.98)	0.98 (0.98, 0.99)	0.95 (0.94, 0.96)	0.95 (0.94, 0.96)	0.94 (0.93, 0.95)
	Precision	0.98 (0.97, 0.99)	0.99 (0.98, 0.99)	0.96 (0.95, 0.97)	0.96 (0.95, 0.97)	0.95 (0.93, 0.96)
	Recall	0.97 (0.96, 0.98)	0.98 (0.97, 0.98)	0.94 (0.93, 0.95)	0.95 (0.93, 0.96)	0.94 (0.93, 0.96)
Age	AUROC	0.94 (0.93, 0.95)	0.94 (0.93, 0.95)	0.91 (0.89, 0.92)	0.90 (0.89, 0.91)	0.84 (0.83, 0.86)
	Accuracy	0.86 (0.85, 0.88)	0.87 (0.85, 0.88)	0.83 (0.81, 0.84)	0.81 (0.80, 0.83)	0.76 (0.74, 0.78)
	Precision	0.84 (0.82, 0.86)	0.84 (0.82, 0.86)	0.81 (0.79, 0.83)	0.80 (0.78, 0.82)	0.75 (0.73, 0.77)
	Recall	0.90 (0.88, 0.91)	0.90 (0.88, 0.92)	0.86 (0.84, 0.87)	0.84 (0.82, 0.86)	0.78 (0.76, 0.81)
b. Masked image						
		VGGFace (aligned)	VGGFace (unaligned)	ImageNet (unaligned)	Places-365 (unaligned)	Flattened-image (unaligned)
Gender	AUROC	0.78 (0.77, 0.80)	0.76 (0.74, 0.77)	0.83 (0.81, 0.85)	0.81 (0.79, 0.82)	0.59 (0.57, 0.61)
	Accuracy	0.72 (0.70, 0.74)	0.70 (0.68, 0.72)	0.77 (0.75, 0.79)	0.75 (0.73, 0.76)	0.58 (0.56, 0.60)
	Precision	0.72 (0.70, 0.75)	0.70 (0.68, 0.73)	0.78 (0.75, 0.80)	0.74 (0.72, 0.77)	0.57 (0.55, 0.60)
	Recall	0.73 (0.70, 0.76)	0.69 (0.66, 0.71)	0.77 (0.74, 0.79)	0.75 (0.73, 0.78)	0.62 (0.59, 0.65)
Age	AUROC	0.63 (0.61, 0.65)	0.63 (0.60, 0.65)	0.64 (0.62, 0.67)	0.64 (0.62, 0.66)	0.53 (0.50, 0.55)
	Accuracy	0.61 (0.58, 0.62)	0.60 (0.57, 0.61)	0.61 (0.59, 0.63)	0.62 (0.60, 0.63)	0.53 (0.51, 0.54)
	Precision	0.60 (0.57, 0.62)	0.59 (0.56, 0.61)	0.60 (0.57, 0.63)	0.61 (0.58, 0.64)	0.53 (0.50, 0.56)
	Recall	0.64 (0.61, 0.67)	0.64 (0.62, 0.67)	0.63 (0.60, 0.65)	0.63 (0.60, 0.66)	0.52 (0.49, 0.54)

Table 3

Performance of gender, age and race classification using UTK-Face sample

a. Original image						
		VGGFace (aligned)	VGGFace (unaligned)	ImageNet (unaligned)	Places-365 (unaligned)	Flattened-image (unaligned)
Gender	AUROC	0.98 (0.98, 0.98)	0.97 (0.96, 0.97)	0.93 (0.93, 0.94)	0.90 (0.90, 0.91)	0.70 (0.69, 0.70)
	Accuracy	0.93 (0.93, 0.93)	0.90 (0.89, 0.90)	0.86 (0.85, 0.86)	0.82 (0.82, 0.83)	0.65 (0.64, 0.65)
	Precision	0.94 (0.93, 0.94)	0.90 (0.90, 0.91)	0.86 (0.86, 0.87)	0.83 (0.82, 0.84)	0.65 (0.64, 0.66)
	Recall	0.93 (0.92, 0.93)	0.89 (0.88, 0.89)	0.85 (0.84, 0.86)	0.82 (0.81, 0.82)	0.64 (0.63, 0.64)
Age	AUROC	0.94 (0.93, 0.94)	0.92 (0.91, 0.92)	0.86 (0.86, 0.87)	0.83 (0.82, 0.83)	0.68 (0.67, 0.68)
	Accuracy	0.85 (0.85, 0.86)	0.83 (0.82, 0.83)	0.77 (0.77, 0.78)	0.75 (0.74, 0.75)	0.63 (0.62, 0.64)
	Precision	0.85 (0.85, 0.86)	0.84 (0.83, 0.84)	0.77 (0.76, 0.78)	0.74 (0.74, 0.75)	0.63 (0.62, 0.64)
	Recall	0.85 (0.84, 0.86)	0.82 (0.81, 0.82)	0.77 (0.76, 0.78)	0.75 (0.74, 0.76)	0.63 (0.62, 0.64)
Race	AUROC	0.96 (0.95, 0.96)	0.93 (0.93, 0.94)	0.86 (0.86, 0.87)	0.83 (0.82, 0.83)	0.70 (0.69, 0.71)
	Accuracy	0.90 (0.90, 0.90)	0.86 (0.85, 0.86)	0.78 (0.78, 0.79)	0.75 (0.74, 0.76)	0.65 (0.64, 0.65)
	Precision	0.90 (0.89, 0.90)	0.85 (0.84, 0.86)	0.79 (0.78, 0.80)	0.76 (0.75, 0.76)	0.65 (0.64, 0.66)
	Recall	0.90 (0.90, 0.91)	0.87 (0.87, 0.88)	0.78 (0.77, 0.78)	0.74 (0.73, 0.75)	0.65 (0.64, 0.66)
b. Masked image						
		VGGFace (aligned)	VGGFace (unaligned)	ImageNet (unaligned)	Places-365 (unaligned)	Flattened-image (unaligned)
Gender	AUROC	0.68 (0.67, 0.68)	0.64 (0.64, 0.65)	0.68 (0.67, 0.69)	0.64 (0.63, 0.65)	0.51 (0.51, 0.52)
	Accuracy	0.63 (0.62, 0.63)	0.61 (0.60, 0.61)	0.64 (0.63, 0.64)	0.60 (0.59, 0.61)	0.51 (0.50, 0.52)
	Precision	0.61 (0.61, 0.62)	0.60 (0.60, 0.61)	0.64 (0.63, 0.65)	0.61 (0.60, 0.62)	0.51 (0.50, 0.52)
	Recall	0.68 (0.67, 0.69)	0.62 (0.61, 0.63)	0.62 (0.61, 0.63)	0.55 (0.54, 0.56)	0.52 (0.51, 0.53)
Age	AUROC	0.63 (0.63, 0.64)	0.61 (0.60, 0.62)	0.65 (0.64, 0.66)	0.63 (0.62, 0.64)	0.56 (0.55, 0.56)
	Accuracy	0.60 (0.59, 0.60)	0.58 (0.57, 0.59)	0.61 (0.60, 0.62)	0.59 (0.59, 0.60)	0.54 (0.53, 0.55)
	Precision	0.59 (0.58, 0.60)	0.58 (0.57, 0.59)	0.61 (0.60, 0.62)	0.59 (0.58, 0.60)	0.54 (0.53, 0.55)
	Recall	0.62 (0.61, 0.63)	0.60 (0.59, 0.61)	0.63 (0.62, 0.63)	0.62 (0.61, 0.62)	0.52 (0.51, 0.53)
Race	AUROC	0.59 (0.58, 0.60)	0.57 (0.57, 0.58)	0.62 (0.61, 0.63)	0.61 (0.60, 0.62)	0.54 (0.53, 0.54)
	Accuracy	0.57 (0.56, 0.57)	0.55 (0.54, 0.56)	0.59 (0.58, 0.59)	0.58 (0.57, 0.58)	0.53 (0.53, 0.54)
	Precision	0.56 (0.55, 0.57)	0.55 (0.54, 0.56)	0.59 (0.58, 0.60)	0.58 (0.57, 0.59)	0.53 (0.52, 0.54)
	Recall	0.59 (0.58, 0.60)	0.56 (0.55, 0.57)	0.57 (0.56, 0.58)	0.56 (0.55, 0.57)	0.51 (0.50, 0.52)

Chapter 3

Table 1

Perspectives of artificial intelligence applications in management

Perspective	Automation	Augmentation	Automation-Augmentation Paradox
Theoretical Assumption	As a wide range of tasks are now computerizable, certain jobs that used to be completed only by humans are now automatable.	As they possess differing capabilities and limitations, humans and machines challenge each other, integrate another's knowledge or collaborate when performing tasks.	As augmentation is not neatly separatable from automation, organizations should adopt a broader perspective comprising both strategies.
Examples of Dependent Variable	The probability tasks, occupations, industries or geographies are computerizable	Collaboration configuration, organizational design	Positive or negative organizational or societal outcomes
Examples of Moderating Variable	-	Opacity in machine's decision processes	Strategies of artificial intelligence applications
Examples of Independent Variable	Tasks, capabilities, skills or occupations	Capabilities or limitations of humans and machines in performing tasks	Tasks, capabilities, skills or occupations
Representative Authors	Frey & Osborne, 2017	Brynjolfsson & McAfee, 2014	Raisch & Krakowski, 2021

Figures

Chapter 1

Figure 1

Comparison of the difference of the means of self-presentational facial attributes by sexual orientation and gender for study 1a

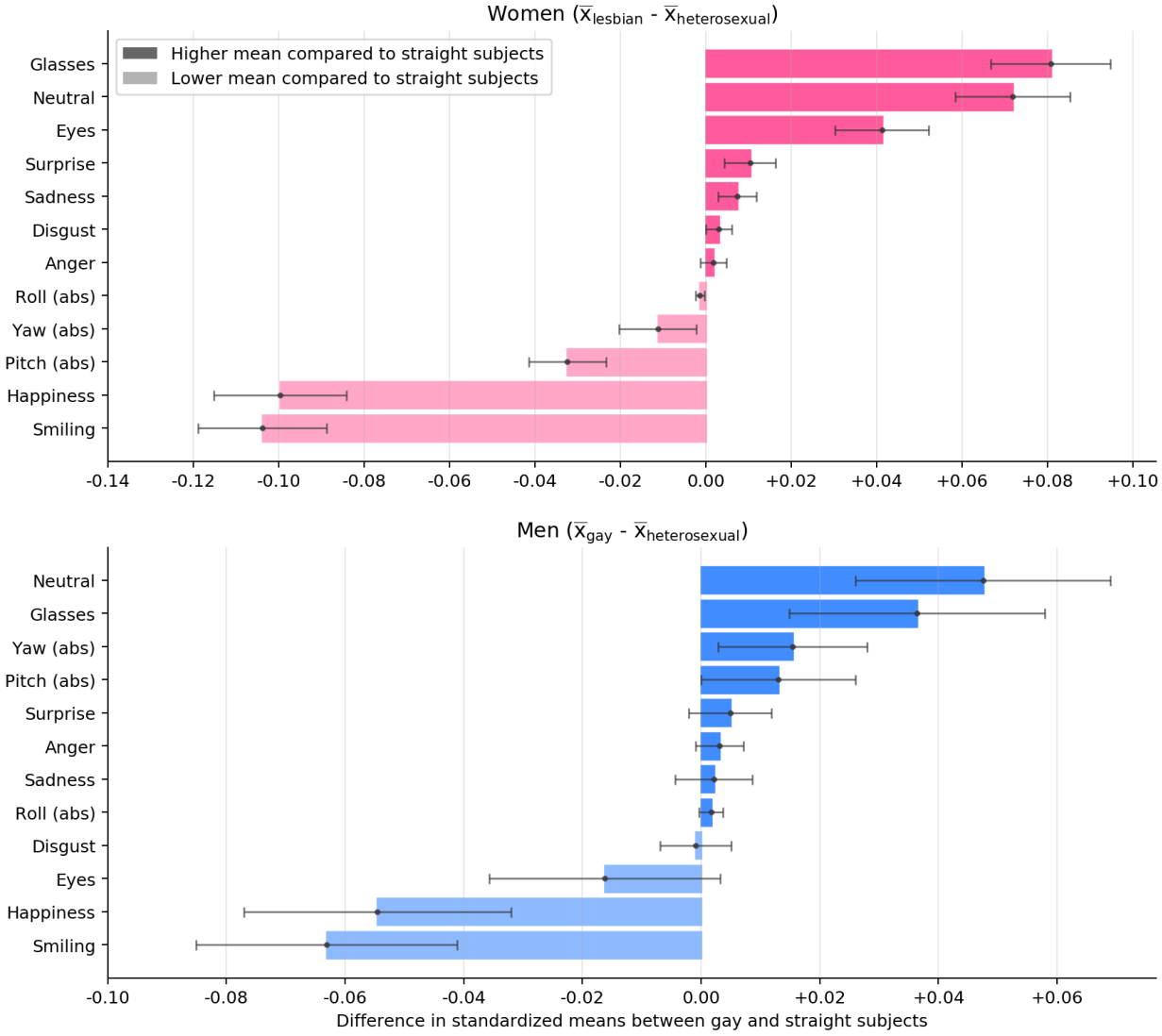


Figure 2

Average AUC afforded by self-presentational facial attributes by gender for study 1b

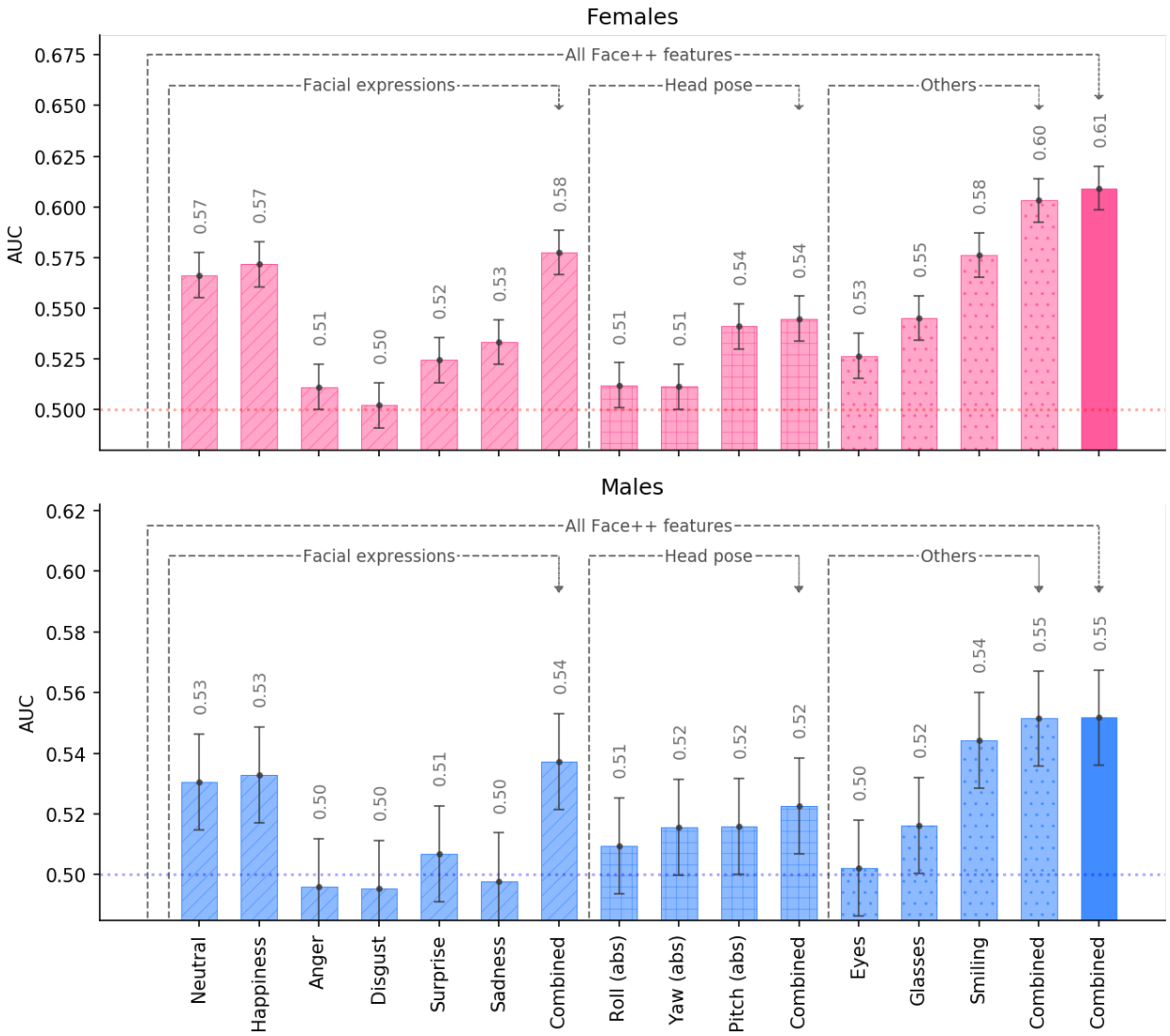


Figure 3

Baseline AUC results by number of images and gender for study 1c

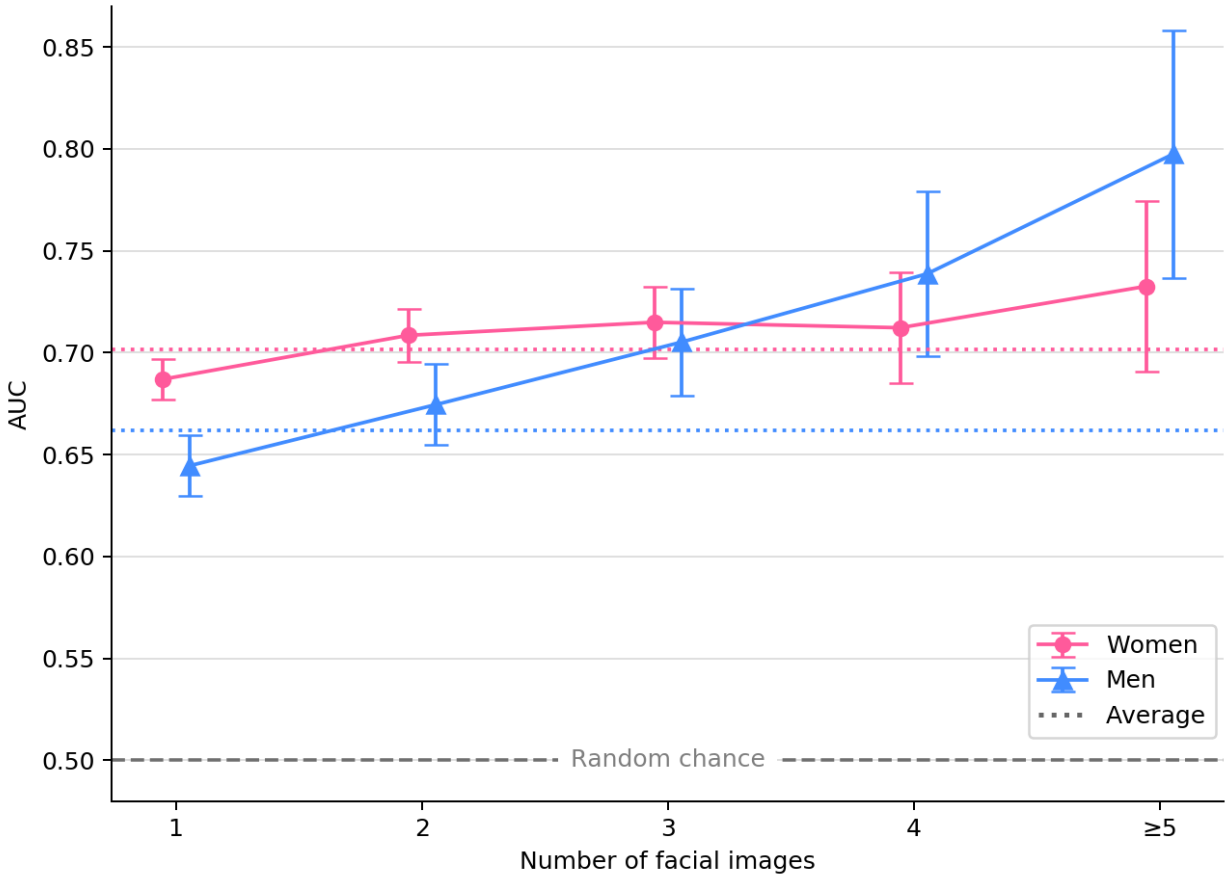
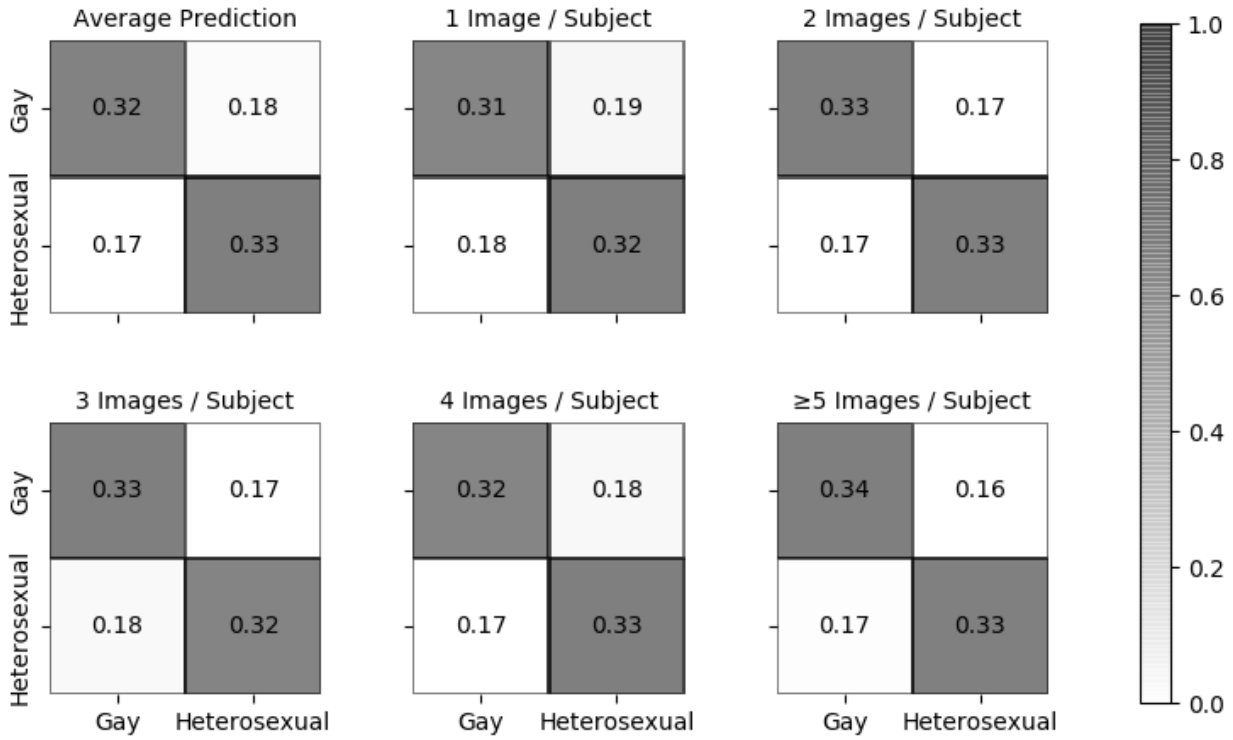


Figure 4

Baseline confusion matrices by number of images and gender for study 1c

Women:



Men:

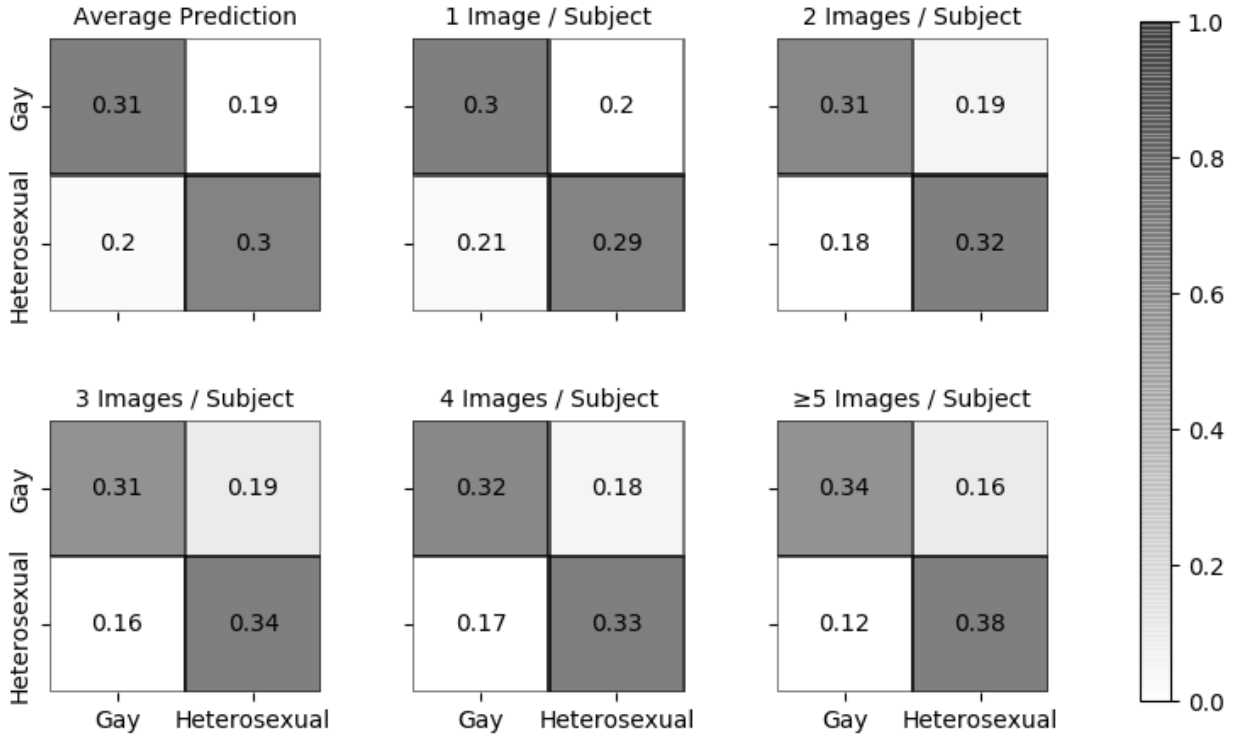
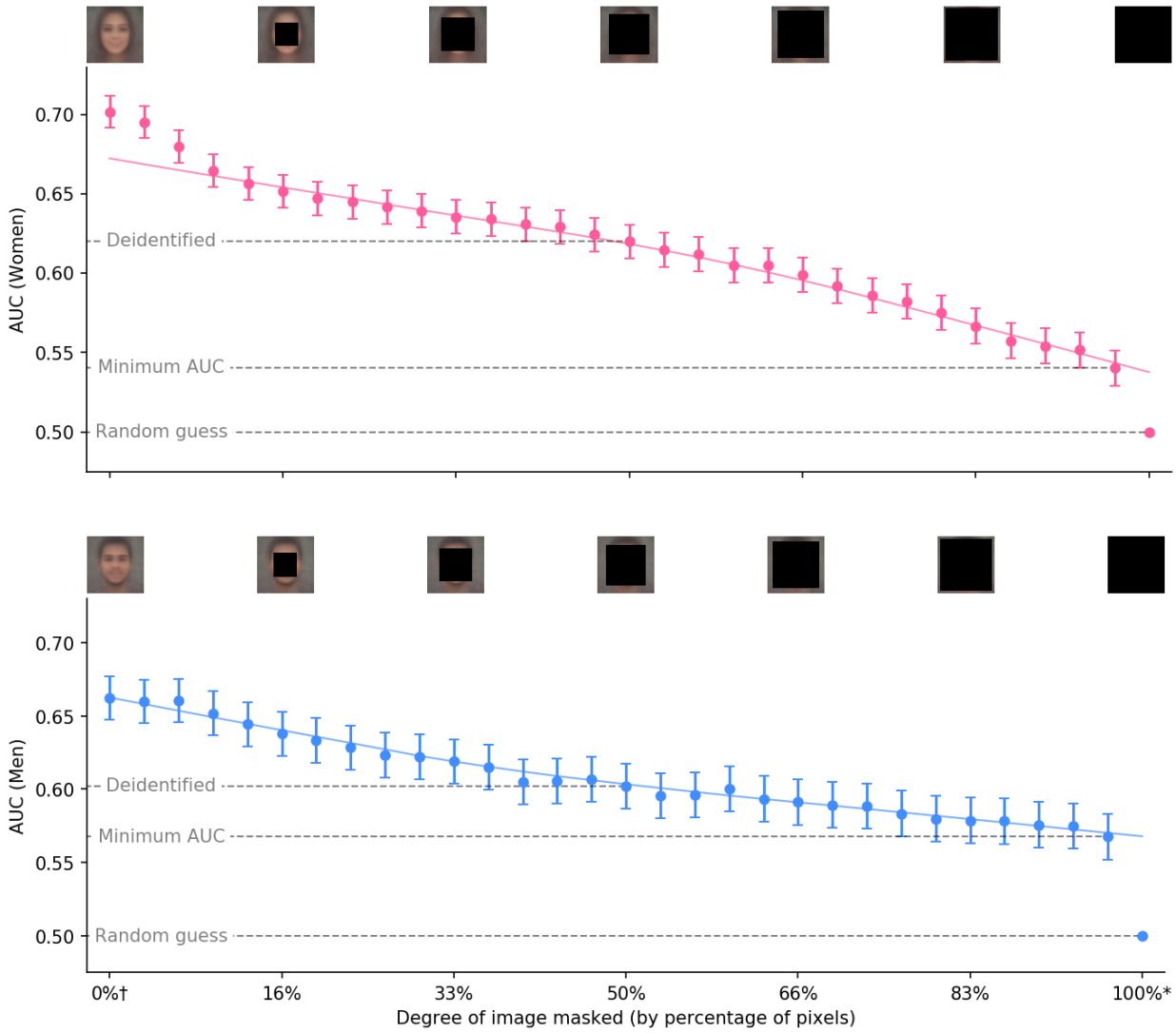

Average Prediction
1 Image / Subject
2 Images / Subject
3 Images / Subject
4 Images / Subject
≥5 Images / Subject
Gay
Heterosexual
0.31
0.19
0.2
0.3
0.3
0.2
0.21
0.29
0.31
0.19
0.18
0.32
0.31
0.19
0.16
0.34
0.32
0.18
0.17
0.33
0.34
0.16
0.12
0.38
1.0
0.8
0.6
0.4
0.2
0.0

Figure 5

AUC results by different degrees of masking compared to auc of random classifications (fully masked images) for study 1c



Note: † baseline AUC. * random AUC.

Figure 6

An example of the facial mask used to separate the facial regions for study 2a

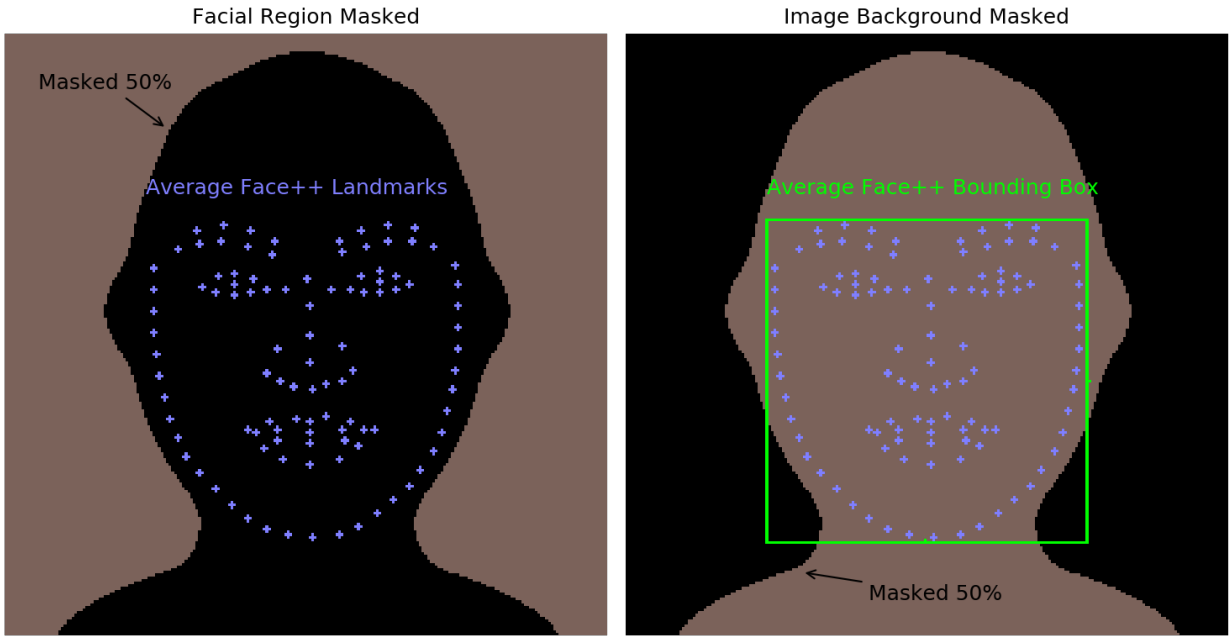


Figure 7

Image brightness by sexual orientation, facial regions, and gender for study 2a

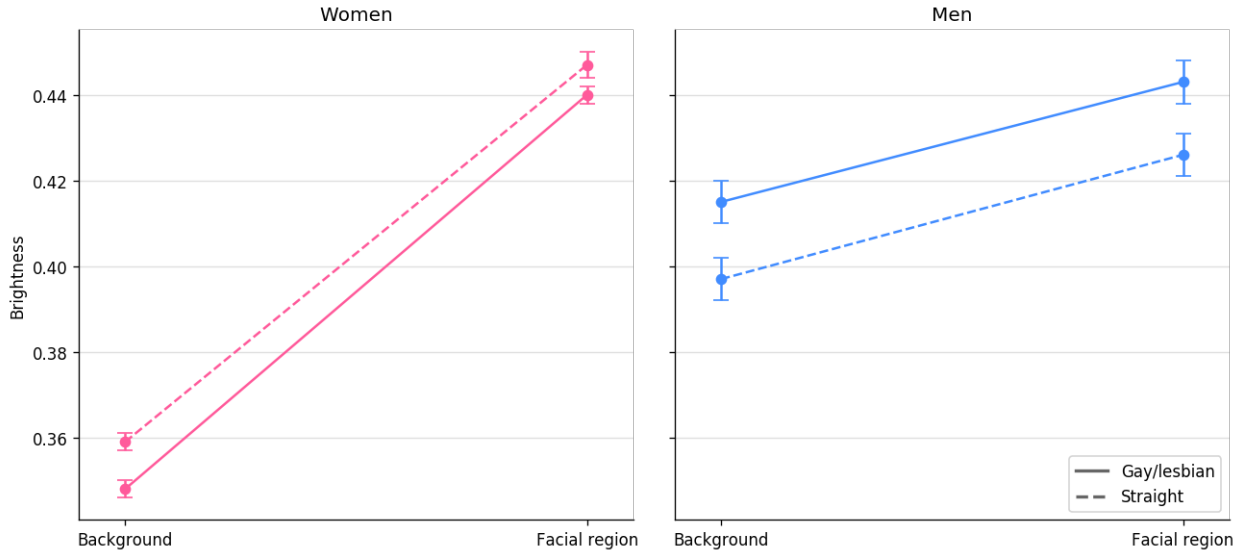
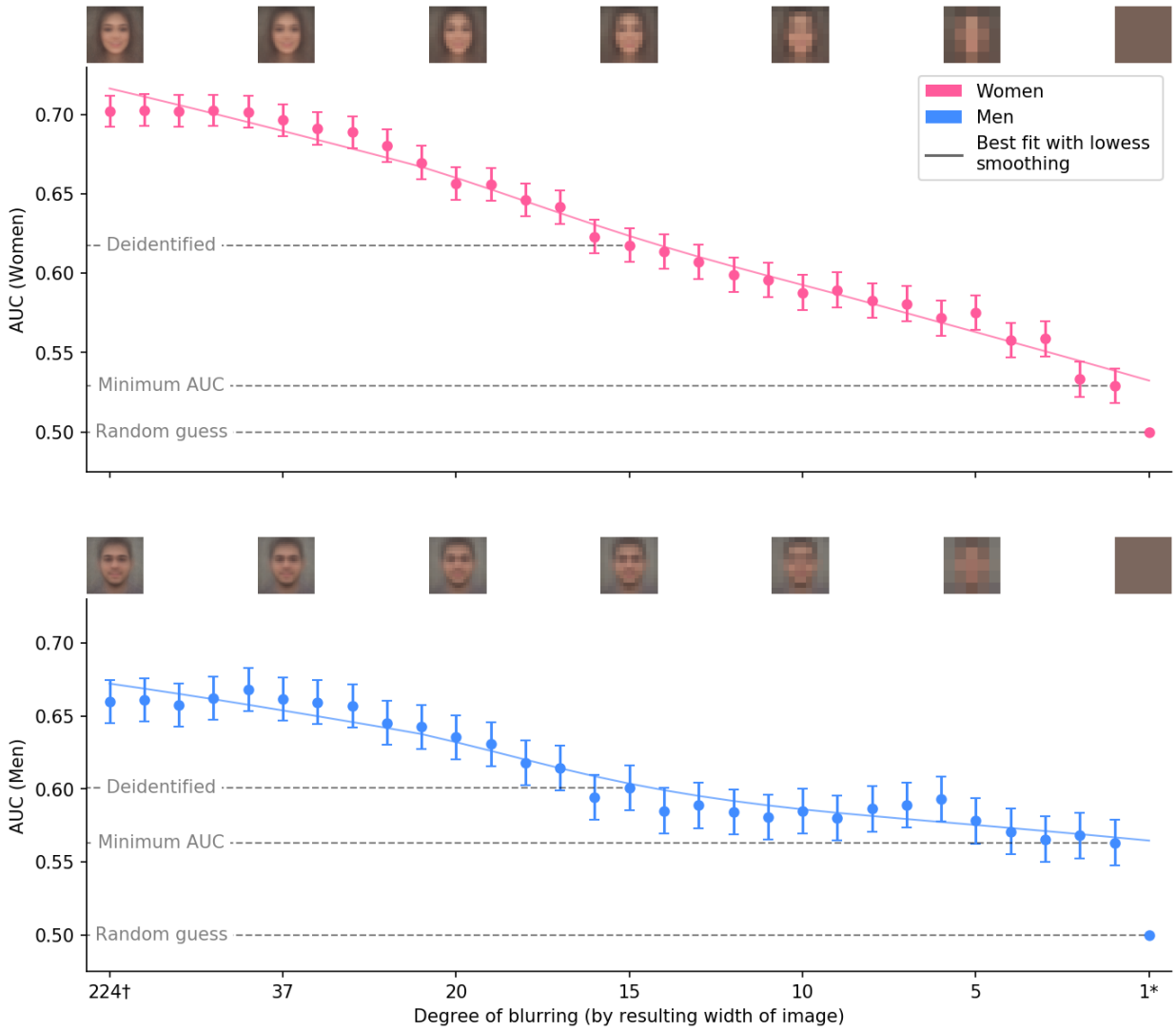


Figure 8

AUC results by different degrees of blurring compared to AUC of random (one-pixel blurred image of dataset) for study 2b

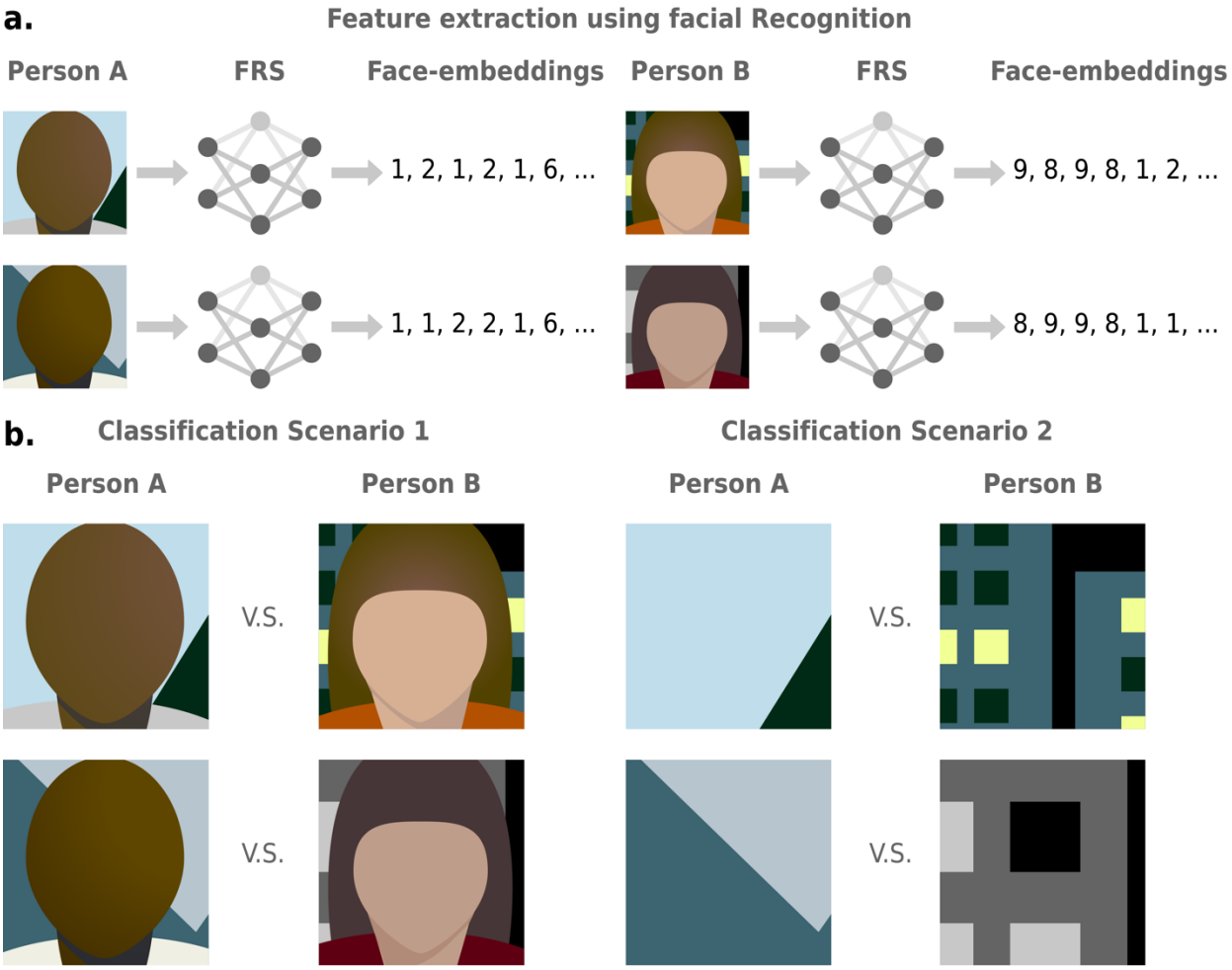


Note: † baseline AUC. * random AUC.

Chapter 2

Figure 1

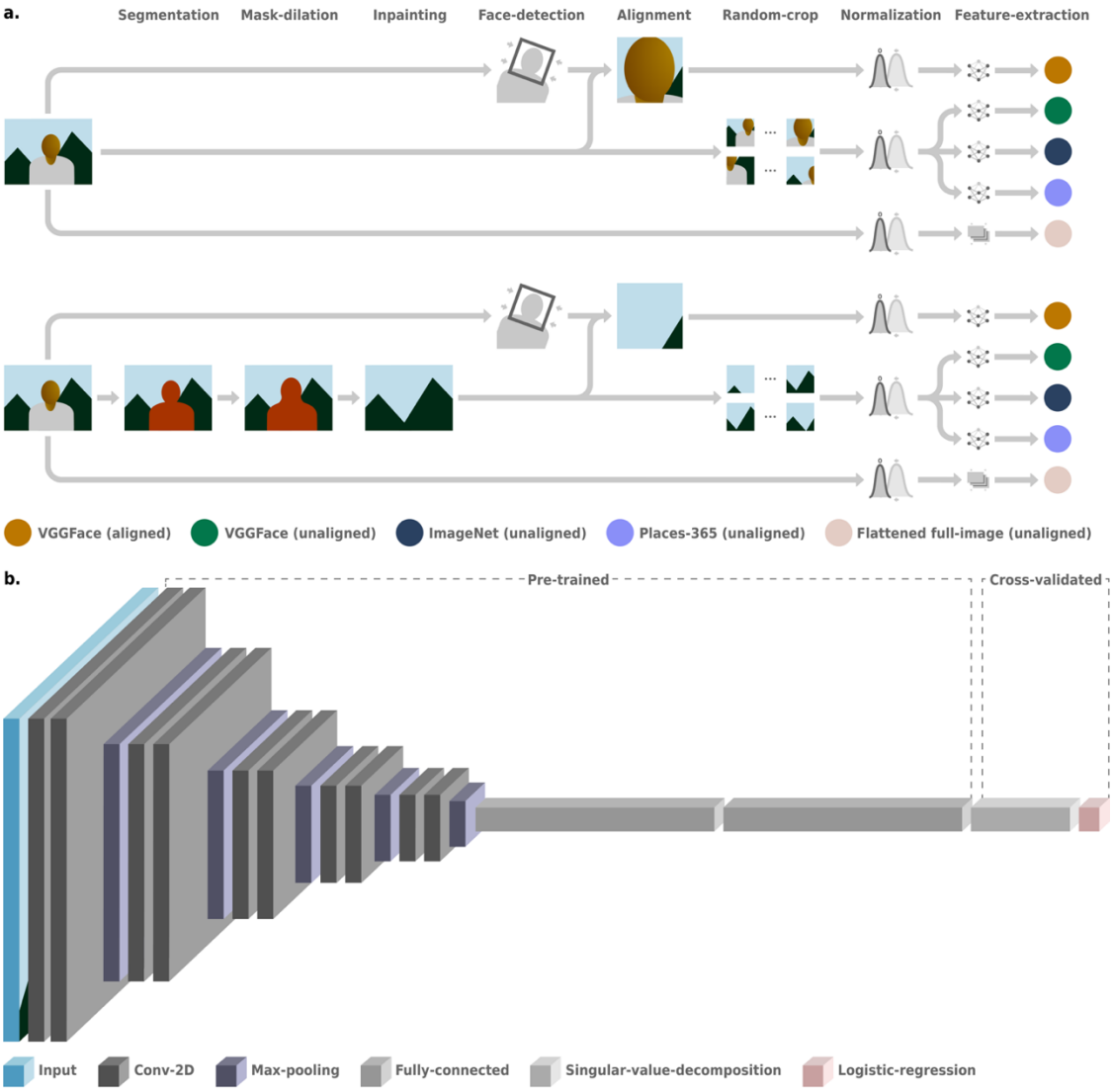
A toy example of a facial recognition system (FRS)



a., Four facial images of two unique individuals, Person A and Person B, were entered into a toy facial recognition model. While the images were different, they produced identical face-embeddings for each person but different numbers between the two people. Companies would normally use the unique face embeddings to further extract other sensitive information such as sexual orientation or political orientation. **b.,** Consider an universe with only two people, Person A and B. Person A enjoys outdoor activities, thus all photographs of Person A were taken during the day in rural places. Person B enjoys nightlife, so all Person B's photographs were taken at night in urban places. In Classification Scenario 1, the facial recognition system successfully classified the two people (such as their identity or sensitive traits) but it is unclear whether the background and lighting contributed to the classifications. In Classification Scenario 2, faces were removed, but companies profiling these individuals were still able to classify identity or traits of Person A and Person B using these images because of their consistent behavioral tendencies.

Figure 2

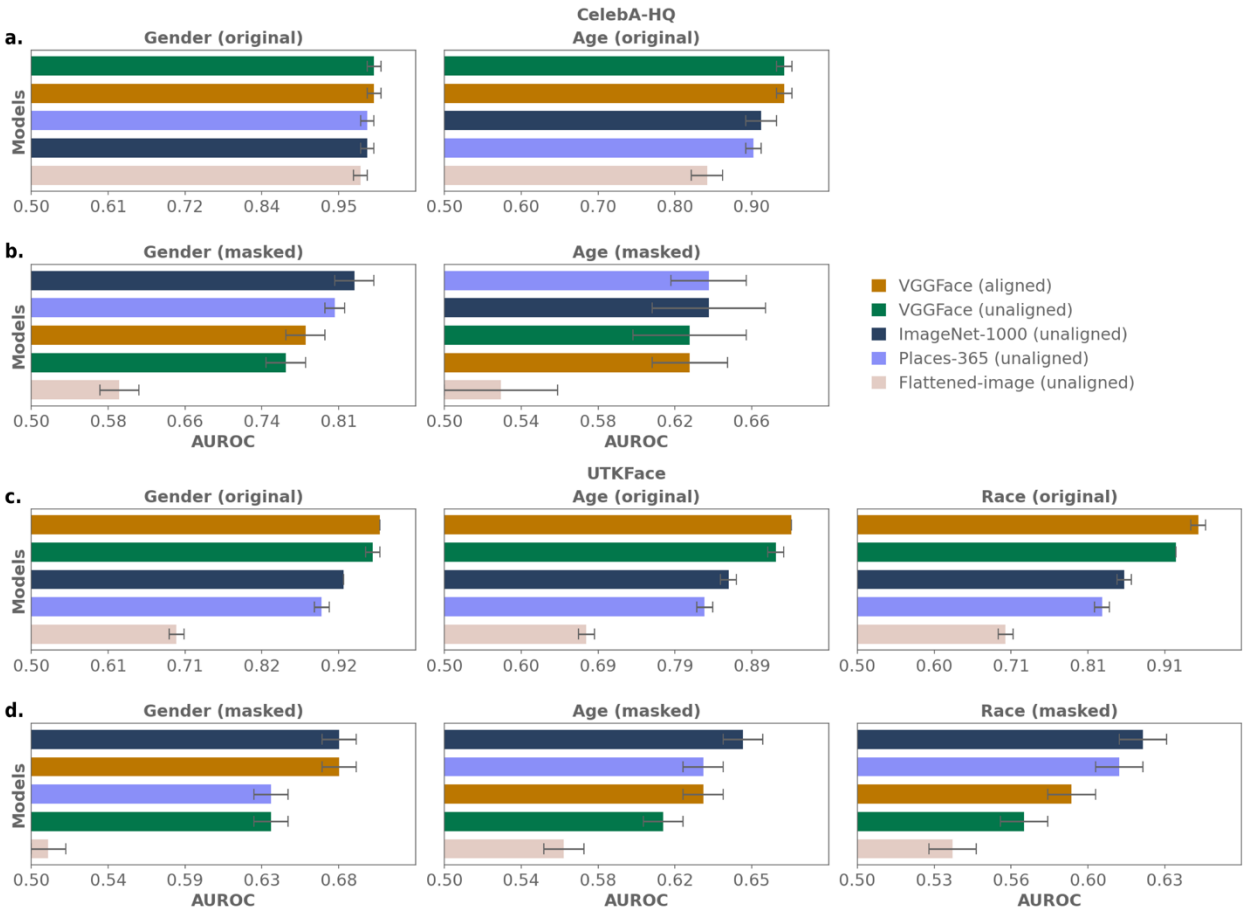
Overview of image preprocessing and training pipelines



a., In the upper half of the figure, raw photographs were entered into a face-detector, called MTCNN (K. Zhang et al., 2016). We used the left and right eye positions as anchor points, to which we rotated the photograph so that the result would contain eye positions at the same target locations. For images not aligned, we created a total of 30 random crops around each photograph, varying in sizes and positions, plus the original uncropped photograph. The cropped and original photograph were separately entered into the face, image and scene recognition model to create 31 sets of embeddings for each model. The embeddings were then averaged to produce one set of embedding per model. For baseline, the image was simply flattened into a single array. All images were resized to 224 by 224 pixels and normalized to a range of -1 to 1 before feature extraction. In lower half of the figure, we employed selfie-segmentation using Google's MediaPipe (Lugaresi et al., 2019). We dilated the segmentation mask by 5% of the image width to remove edge information. We then filled the mask using inpainting technique (Bertalmio et al., 2001). **b.**, An example of the VGG-16 deep neural network model and custom-training process is shown. Only the 4,096 scores from the final embedding layer of each deep neural network model is extracted. The original labels (not shown in figure) were not used. The extracted embedding scores were then cross-validated to produce the sexual orientation predictions.

Figure 3

Comparison of classification performance



a., Performance of demographic classification using original photographs from CelebA-HQ dataset (Guo et al., 2016; Karras et al., 2018), ranked from the model with the highest performance to the lowest. **b.**, Performance using de-identified (masked) photographs from CelebA-HQ dataset. **c**., Performance of demographic classification using original photographs from UTKFace dataset (Z. Zhang et al., 2017). **d**., Performance using de-identified (masked) photographs from UTKFace dataset. Confidence intervals shown in grey.

References

Introduction

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of Personality and Social Psychology, 64(3), 431–441. https://doi.org/10.1037/0022-3514.64.3.431

Aravinda, T., Krishnareddy, K., Varghese, S., Chandrika, P. V., Rao, T. P., & Trofimov, V. (2022). Implementation of Facial Recognition (AI) and Its Impact on the Service Sector. 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 74–80. https://doi.org/10.1109/ICAAIC53929.2022.9793161

Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixão, T. M., Mutz, F., de Paula Veronese, L., Oliveira-Santos, T., & De Souza, A. F. (2021). Self-driving cars: A survey. Expert Systems with Applications, 165, 113816. https://doi.org/10.1016/j.eswa.2020.113816

Boudette, N. E., Metz, C., & Ewing, J. (2022, June 15). Tesla Autopilot and Other Driver-Assist Systems Linked to Hundreds of Crashes. The New York Times. https://www.nytimes.com/2022/06/15/business/self-driving-car-nhtsa-crash-data.html

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. Strategic Management Journal, 42(1), 30–57. https://doi.org/10.1002/smj.3215

Csaszar, F. A., & Steinberger, T. (2022). Organizations as Artificial Intelligences: The Use of Artificial Intelligence Analogies in Organization Theory. Academy of Management Annals, 16(1), 1–37. https://doi.org/10.5465/annals.2020.0192

Dai, T., Sycara, K., & Zheng, R. (2021). Agent Reasoning in AI-Powered Negotiation. In D. M. Kilgour & C. Eden (Eds.), Handbook of Group Decision and Negotiation (pp. 1187–1211). Springer International Publishing. https://doi.org/10.1007/978-3-030-49629-6_26

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Drozdowski, P., Rathgeb, C., Dantcheva, A., Damer, N., & Busch, C. (2020). Demographic Bias in Biometrics: A Survey on an Emerging Challenge. IEEE Transactions on Technology and Society, 1(2), 89–103. https://doi.org/10.1109/TTS.2020.2992344

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: Beyond Jeopardy! Artificial Intelligence, 199–200, 93–105. https://doi.org/10.1016/j.artint.2012.06.009

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. Nature Machine Intelligence, 2(11), 665–673. https://doi.org/10.1038/s42256-020-00257-z

Harwell, D. (2019, November 6). Rights group files federal complaint against AI-hiring firm HireVue, citing 'unfair and deceptive' practices. Washington Post. https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/

Hu, S., Xiong, J., Fu, P., Qiao, L., Tan, J., Jin, L., & Tang, K. (2017). Signatures of personality on dense 3D facial images. Scientific Reports, 7(1), 73. https://doi.org/10.1038/s41598-017-00071-5

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the big five personality traits using real-life static facial images. Scientific Reports, 10(1), 8487. https://doi.org/10.1038/s41598-020-65358-6

Kosinski, M. (2017). Facial Width-to-Height Ratio Does Not Predict Self-Reported Behavioral Tendencies. Psychological Science, 28(11), 1675–1682. https://doi.org/10.1177/0956797617716929

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. Scientific Reports, 11(1), 100. https://doi.org/10.1038/s41598-020-79310-1

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2020). Ghost in the Machine: On Organizational Theory in the Age of Machine Learning. Academy of Management Review, amr.2019.0247. https://doi.org/10.5465/amr.2019.0247

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading Digits in Natural Images with Unsupervised Feature Learning. NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011. http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf

Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, 103517. https://doi.org/10.1016/j.autcon.2020.103517

Peña, A., Serna, I., Morales, A., & Fierrez, J. (2020). Bias in Multimodal AI: Testbed for Fair Automatic Recruitment. https://arxiv.org/abs/2004.07173v1

Rule, N., & Ambady, N. (2010). First Impressions of the Face: Predicting Success. Social and Personality Psychology Compass, 4(8), 506–516. https://doi.org/10.1111/j.1751-9004.2010.00282.x

Schaller, R. R. (1997). Moore's law: Past, present and future. IEEE Spectrum, 34(6), 52–59. https://doi.org/10.1109/6.591665

Sheetal, A., Feng, Z., & Savani, K. (2020). Using Machine Learning to Generate Novel Hypotheses: Increasing Optimism About COVID-19 Makes People Less Willing to Justify Unethical Behaviors. Psychological Science, 31(10), 1222–1235. https://doi.org/10.1177/0956797620959594

Sheetal, A., & Savani, K. (2021). A Machine Learning Model of Cultural Change: Role of Prosociality, Political Attitudes, and Protestant Work Ethic. 16.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. Nature, 550(7676), 354–359. https://doi.org/10.1038/nature24270

Stoker, J. I., Garretsen, H., & Spreeuwers, L. J. (2016). The Facial Appearance of CEOs: Faces Signal Selection but Not Performance. PLOS ONE, 11(7), e0159950. https://doi.org/10.1371/journal.pone.0159950

Suresh, H., & Guttag, J. V. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. Equity and Access in Algorithms, Mechanisms, and Optimization, 1–9. https://doi.org/10.1145/3465416.3483305

Tiku, N. (2022, June 11). The Google engineer who thinks the company's AI has come to life. Washington Post. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Wang, D. (2021). Presentation in self-posted facial images can expose sexual orientation: Implications for research and privacy. https://doi.org/10.31234/osf.io/u7vcd

Wang, D. (under review). How Existing Biometric Privacy Acts Would Fail at Protecting People's Privacy in Self-Posted Facial Photographs.

Wang, D., Nair, K., Kouchaki, M., Zajac, E. J., & Zhao, X. (2019). A Case of Evolutionary Mismatch? Why Facial Width-to-Height Ratio May Not Predict Behavioral Tendencies. Psychological Science, 30(7), 1074–1081. https://doi.org/10.1177/0956797619849928

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of Personality and Social Psychology, 114(2), 246–257. https://doi.org/10.1037/pspa0000098

Wu, X., & Zhang, X. (2017). Responses to Critiques on Machine Learning of Criminality Perceptions (Addendum of arXiv:1611.04135). ArXiv:1611.04135 [Cs]. http://arxiv.org/abs/1611.04135

Chapter 1

Agüera y Arcas, B., Todorov, A., & Mitchell, M. (2018, January 11). Do algorithms reveal sexual orientation or just expose our stereotypes? Medium. https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest, 20(1), 1–68. https://doi.org/10.1177/1529100619832930

Biecek, P. (2018). DALEX: explainers for complex predictive models in R. The Journal of Machine Learning Research, 19(1), 3245-3249.

Conger, K., Fausset, R., & Kovaleski, S. F. (2019, May 14). San Francisco bans facial recognition technology. New York Times. Retrieved from https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics, 44(3), 837–845. https://doi.org/10.2307/2531595

Gecer, B., Ploumpis, S., Kotsia, I., & Zafeiriou, S. (2019). GANFIT: Generative Adversarial Network Fitting for High Fidelity 3D Face Reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1155–1164. https://openaccess.thecvf.com/content_CVPR_2019/html/Gecer_GANFIT_Generative_Adversarial_Network_Fitting_for_High_Fidelity_3D_Face_CVPR_2019_paper.html

Goffman, E. (1959). The presentation of self in everyday life. Anchor.

Haferkamp, N., Eimler, S. C., Papadakis, A.-M., & Kruck, J. V. (2012). Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites. Cyberpsychology, Behavior, and Social Networking, 15(2), 91–98. https://doi.org/10.1089/cyber.2011.0151

Hancock, J. T., & Toma, C. L. (2009). Putting your best face forward: The accuracy of online dating photographs. Journal of Communication, 59(2), 367–386. https://doi.org/10.1111/j.1460-2466.2009.01420.x

Jaeger, B., Sleegers, W. W. A., & Evans, A. M. (2020). Automated classification of demographics from face images: A tutorial and validation. Social and Personality Psychology Compass, 14(3), e12520. https://doi.org/10.1111/spc3.12520

Jenkins, R., & Burton, A. M. (2011). Stable face representations. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1571), 1671–1683. https://doi.org/10.1098/rstb.2010.0379

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

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the Big Five personality traits using real-life static facial images. Scientific Reports, 10(1), 8487. https://doi.org/10.1038/s41598-020-65358-6

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. Cognition & Emotion, 13(5), 505–521. https://doi.org/10.1080/026999399379168

Kittler, J., Huber, P., Feng, Z.-H., Hu, G., & Christmas, W. (2016). 3D morphable face models and their applications. In F. J. Perales & J. Kittler (Eds.), Articulated motion and deformable objects (pp. 185–206). Springer International Publishing. https://doi.org/10.1007/978-3-319-41778-3_19

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. Scientific Reports, 11(1), 100. https://doi.org/10.1038/s41598-020-79310-1

Leary, M. R., & Allen, A. B. (2011). Personality and persona: Personality processes in self-presentation. Journal of Personality, 79(6), 1191–1218. https://doi.org/10.1111/j.1467-6494.2010.00704.x

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. Psychological Bulletin 107(1), 34–47.

Leary, M. R., Nezlek, J. B., Downs, D., Radford-Davenport, J., Martin, J., & McMullen, A. (1994). Self-presentation in everyday interactions: Effects of target familiarity and gender composition. Journal of Personality and Social Psychology, 67(4), 664–673. https://doi.org/10.1037/0022-3514.67.4.664

Li, Y., Vishwamitra, N., Hu, H., Knijnenburg, B. P., & Caine, K. (2017). Effectiveness and users' experience of face blurring as a privacy protection for sharing photos via online social networks. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61(1), 803–807. https://doi.org/10.1177/1541931213601694

Matz, S. C., Appel, R. E., & Kosinski, M. (2020). Privacy in the age of psychological targeting. Current Opinion in Psychology, 31, 116–121. https://doi.org/10.1016/j.copsyc.2019.08.010

Matz, S. C., & Harari, G. M. (2021). Personality–place transactions: Mapping the relationships between Big Five personality traits, states, and daily places. Journal of Personality and Social Psychology, 120(5), 1367–1385. https://doi.org/10.1037/pspp0000297

Matzner, T., Masur, P. K., Ochs, C., & von Pape, T. (2016). Do-it-yourself data protection—empowerment or burden? In S. Gutwirth, R. Leenes, & P. De Hert (Eds.), Data protection on the move: Current developments in ICT and privacy/data protection (pp. 277–305). Springer Netherlands. https://doi.org/10.1007/978-94-017-7376-8_11

Nicholls, M. E. R., Wolfgang, B. J., Clode, D., & Lindell, A. K. (2002). The effect of left and right poses on the expression of facial emotion. Neuropsychologia, 40(10), 1662–1665. https://doi.org/10.1016/S0028-3932(02)00024-6

Noyes, E., & Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. Cognition, 165, 97–104. https://doi.org/10.1016/j.cognition.2017.05.012

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. Trends in Cognitive Sciences, 18(11), 566–570. https://doi.org/10.1016/j.tics.2014.09.007

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. Journal of Experimental Social Psychology, 46(2), 315–324. https://doi.org/10.1016/j.jesp.2009.12.002

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oosterhof, N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. Emotion, 9(1), 128–133. https://doi.org/10.1037/a0014520

Ota, C., & Nakano, T. (2021). Neural correlates of beauty retouching to enhance attractiveness of self-depictions in women. Social Neuroscience, 16(2), 121–133. https://doi.org/10.1080/17470919.2021.1873178

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. Proceedings of the British Machine Vision Conference 2015, 41.1-41.12. https://doi.org/10.5244/C.29.41

Phillips, P. J. (2017). A cross benchmark assessment of a deep convolutional neural network for face recognition. 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), 705–710. https://doi.org/10.1109/FG.2017.89

Prasad, P., Pathak, R., Gunjan, V. K., & Rao, H. V. (2020). Deep learning based representation for face recognition. Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering (pp. 419–424). https://doi.org/10.1007/978-981-13-8715-9_50

Raschka, S. (2020). Model evaluation, model selection, and algorithm selection in machine learning. ArXiv:1811.12808 [Cs, Stat]. http://arxiv.org/abs/1811.12808

Reddy, V. (2000). Coyness in early infancy. Developmental Science, 3(2), 186–192. https://doi.org/10.1111/1467-7687.00112

Rezende, I. N. (2020). Facial recognition in police hands: Assessing the 'Clearview case' from a European perspective. New Journal of European Criminal Law, 11(3), 375–389. https://doi.org/10.1177/2032284420948161

Rothstein, M. A., & Tovino, S. A. (2019). California takes the lead on data privacy law. Hastings Center Report, 49(5), 4–5. https://doi.org/10.1002/hast.1042

Rudd, N. A. (1996). Appearance and self-presentation research in gay consumer cultures: Issues and impact. Journal of Homosexuality, 31(1–2), 109–134. https://doi.org/10.1300/J082v31n01_07

Sawilowsky, S. S. (2009). New effect size rules of thumb. Journal of Modern Applied Statistical Methods, 8(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Schau, H. J., & Gilly, M. C. (2003). We are what we post? Self-presentation in personal web space. Journal of Consumer Research, 30(3), 385–404. https://doi.org/10.1086/378616

Schlenker, B. R. (2012). Self-presentation. In M. R. Leary & J. P. Tangney (Eds.), Handbook of self and identity (2nd ed., pp. 542–570). The Guilford Press.

Serengil, S. I., & Ozpinar, A. (2020). LightFace: A hybrid deep face recognition framework. 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 1–5. https://doi.org/10.1109/ASYU50717.2020.9259802

Shan, S., Wenger, E., Zhang, J., Li, H., Zheng, H., & Zhao, B. Y. (2020). Fawkes: Protecting privacy against unauthorized deep learning models. 1589–1604. https://www.usenix.org/conference/usenixsecurity20/presentation/shan

Shorten, C., & Khoshgoftaar, T. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6. https://doi.org/10.1186/s40537-019-0197-0

Stoker, J. I., Garretsen, H., & Spreeuwers, L. J. (2016). The facial appearance of CEOs: faces signal selection but not performance. PLOS ONE, 11(7), e0159950. https://doi.org/10.1371/journal.pone.0159950

Susan, S., & Kumar, A. (2021). The balancing trick: Optimized sampling of imbalanced datasets—A brief survey of the recent State of the Art. Engineering Reports, 3(4), e12298. https://doi.org/10.1002/eng2.12298

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2015). Web-scale training for face identification. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2746–2754. https://doi.org/10.1109/CVPR.2015.7298891

Todorov, A. (2008). Evaluating faces on trustworthiness. Annals of the New York Academy of Sciences, 1124(1), 208–224. https://doi.org/10.1196/annals.1440.012

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. Social Cognitive and Affective Neuroscience, 3(2), 119–127. https://doi.org/10.1093/scan/nsn009

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(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. Psychological Science, 25(7), 1404–1417. https://doi.org/10.1177/0956797614532474

Tong, S. T., Corriero, E. F., Wibowo, K. A., Makki, T. W., & Slatcher, R. B. (2020). Self-presentation and impressions of personality through text-based online dating profiles: A lens model analysis. New Media & Society, 22(5), 875–895. https://doi.org/10.1177/1461444819872678

Torralba, A., Fergus, R., & Freeman, W. T. (2008). 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11), 1958–1970. https://doi.org/10.1109/TPAMI.2008.128

Tran, A. T., Hassner, T., Masi, I., Paz, E., Nirkin, Y., & Medioni, G. (2018). Extreme 3D face reconstruction: Seeing through occlusions. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3935–3944. https://doi.org/10.1109/CVPR.2018.00414

Wang, D., Nair, K., Kouchaki, M., Zajac, E. J., & Zhao, X. (2019). A case of evolutionary mismatch? Why facial width-to-height ratio may not predict behavioral tendencies. Psychological Science, 30(7), 1074–1081. https://doi.org/10.1177/0956797619849928

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of Personality and Social Psychology, 114(2), 246–257. https://doi.org/10.1037/pspa0000098

White, D., Sutherland, C. A. M., & Burton, A. L. (2017). Choosing face: The curse of self in profile image selection. Cognitive Research: Principles and Implications, 2(1), 23. https://doi.org/10.1186/s41235-017-0058-3

Witkower, Z., & Tracy, J. L. (2019). A facial-action imposter: How head tilt influences perceptions of dominance from a neutral face. Psychological Science, 30(6), 893–906. https://doi.org/10.1177/0956797619838762

Wolffhechel, K., Fagertun, J., Jacobsen, U. P., Majewski, W., Hemmingsen, A. S., Larsen, C. L., Lorentzen, S. K., & Jarmer, H. (2014). Interpretation of appearance: The effect of facial features on first impressions and personality. PLOS ONE, 9(9), e107721. https://doi.org/10.1371/journal.pone.0107721

Wrzus, C., Wagner, G. G., & Riediger, M. (2016). Personality-situation transactions from adolescence to old age. Journal of Personality and Social Psychology, 110(5), 782–799. https://doi.org/10.1037/pspp0000054

Wu, X., & Zhang, X. (2016). Automated inference on criminality using face images. arXiv preprint arXiv:1611.04135, 4038-4052. http://arxiv.org/abs/1611.04135

Zeiler, M. D., & Fergus, R. (2013). Visualizing and understanding convolutional networks. ArXiv:1311.2901 [Cs]. http://arxiv.org/abs/1311.2901

Zhang, Y., Lu, Y., Nagahara, H., & Taniguchi, R. (2014). Anonymous camera for privacy protection. 22nd International Conference on Pattern Recognition, 4170–4175. https://doi.org/10.1109/ICPR.2014.715

Chapter 2

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age. Journal of Consumer Psychology, 30(4), 736–758. https://doi.org/10.1002/jcpy.1191

Agüera y Arcas, B., Todorov, A., & Mitchell, M. (2018, January 18). Do algorithms reveal sexual orientation or just expose our stereotypes? Medium. https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477

Almeida, D., Shmarko, K., & Lomas, E. (2021). The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: A comparative analysis of US, EU, and UK regulatory frameworks. AI and Ethics. https://doi.org/10.1007/s43681-021-00077-w

BCLP Law. (2022). US biometric laws & pending legislation tracker. https://www.bclplaw.com/images/content/3/2/v2/320807/BIPA-Tracker-II-603732145.3.pdf

Bertalmio, M., Bertozzi, A. L., & Sapiro, G. (2001). Navier-stokes, fluid dynamics, and image and video inpainting. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1, I-355-I–362. https://doi.org/10.1109/CVPR.2001.990497

Biometric Information Privacy Act, Illinois Compiled Statutes § 740.14 (2008). https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57

California Consumer Privacy Act (CCPA). (2018, October 15). State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

Cao, D., Cunjian Chen, Piccirilli, M., Adjeroh, D., Bourlai, T., & Ross, A. (2011). Can facial metrology predict gender? 2011 International Joint Conference on Biometrics (IJCB), 1–8. https://doi.org/10.1109/IJCB.2011.6117471

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. ArXiv:1710.08092 [Cs]. http://arxiv.org/abs/1710.08092

Crawford, N. C., Graves, L., & Katzenstein, J. (2021). Racial Profiling and Islamophobia. The Costs of War. https://watson.brown.edu/costsofwar/costs/social/rights/profiling

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Ebert, T., Götz, F. M., Gladstone, J. J., Müller, S. R., & Matz, S. C. (2021). Spending reflects not only who we are but also who we are around: The joint effects of individual and geographic personality on consumption. Journal of Personality and Social Psychology, 121(2), 378.

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. Nature Machine Intelligence, 2(11), 665–673. https://doi.org/10.1038/s42256-020-00257-z

Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), Computer Vision – ECCV 2016 (pp. 87–102). Springer International Publishing. https://doi.org/10.1007/978-3-319-46487-9_6

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data—A taxonomy of data-driven business models used by start-up firms. International Journal of Operations & Production Management, 36(10), 1382–1406. https://doi.org/10.1108/IJOPM-02-2014-0098

Helveston, M. N. (2018). Reining in commercial exploitation of consumer data symposium. Penn State Law Review, 123(3), 667–702.

Hill, K., & Mac, R. (2021, November 2). Facebook, citing societal concerns, plans to shut down facial recognition system. The New York Times. https://www.nytimes.com/2021/11/02/technology/facebook-facial-recognition.html

Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. (2020). Assessing the big five personality traits using real-life static facial images. Scientific Reports, 10(1), 8487. https://doi.org/10.1038/s41598-020-65358-6

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. ArXiv:1710.10196 [Cs, Stat]. http://arxiv.org/abs/1710.10196

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. Scientific Reports, 11(1), 100. https://doi.org/10.1038/s41598-020-79310-1

Leong, B. (2019). Facial recognition and the future of privacy: I always feel like … somebody's watching me. Bulletin of the Atomic Scientists, 75(3), 109–115. https://doi.org/10.1080/00963402.2019.1604886

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. ArXiv:1906.08172 [Cs]. http://arxiv.org/abs/1906.08172

Matz, S. C., Appel, R. E., & Kosinski, M. (2020). Privacy in the age of psychological targeting. Current Opinion in Psychology, 31, 116–121. https://doi.org/10.1016/j.copsyc.2019.08.010

Matz, S. C., & Harari, G. M. (2020). Personality–place transactions: Mapping the relationships between big-five personality traits, states, and daily places. Journal of Personality and Social Psychology. https://doi.org/10.1037/pspp0000297

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences of the United States of America, 114(48), 12714–12719.

Norberg v. Shutterfly, (United States District Court for the Northern District of Illinois 2015). https://digitalcommons.law.scu.edu/historical/986/

Oosterhof, N. N., & Todorov, A. (2008a). The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2008b). The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Palmer, L. F., Annie. (2021, June 12). Rules around facial recognition and policing remain blurry. CNBC. https://www.cnbc.com/2021/06/12/a-year-later-tech-companies-calls-to-regulate-facial-recognition-met-with-little-progress.html

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. Procedings of the British Machine Vision Conference 2015, 41.1-41.12. https://doi.org/10.5244/C.29.41

Phillips, P. J. (2017). A cross benchmark assessment of a deep convolutional neural network for face recognition. 2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), 705–710. https://doi.org/10.1109/FG.2017.89

Prasad, P., Gunjan, V., & Rao, H. (2020). Deep learning based representation for face recognition. In ICCCE 2019 (pp. 419–424). https://doi.org/10.1007/978-981-13-8715-9_50

Priadana, A., Maarif, M. R., & Habibi, M. (2020). Gender Prediction for Instagram User Profiling using Deep Learning. 2020 International Conference on Decision Aid Sciences and Application (DASA), 432–436. https://doi.org/10.1109/DASA51403.2020.9317143

Ranjan, R., Sankaranarayanan, S., Bansal, A., Bodla, N., Chen, J.-C., Patel, V. M., Castillo, C. D., & Chellappa, R. (2018). Deep learning for understanding faces: Machines may be just as good, or better, than humans. IEEE Signal Processing Magazine, 35(1), 66–83. https://doi.org/10.1109/MSP.2017.2764116

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Santow, E. (2020). Emerging from AI utopia. Science, 368(6486). https://www.science.org/doi/full/10.1126/science.abb9369

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Schwartz, O. (2019, March 6). Don't look now: Why you should be worried about machines reading your emotions. The Guardian. https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science

Stempel, J. (2019, August 8). Facebook loses facial recognition appeal, must face privacy class action. Reuters. https://www.reuters.com/article/us-facebook-privacy-lawsuit-idUSKCN1UY2BZ

Stirrat, M., & Perrett, D. I. (2010). Valid Facial Cues to Cooperation and Trust: Male Facial Width and Trustworthiness. Psychological Science, 21(3), 349–354. https://doi.org/10.1177/0956797610362647

Sundararajan, K., & Woodard, D. L. (2018). Deep Learning for Biometrics: A Survey. ACM Computing Surveys, 51(3), 1–34. https://doi.org/10.1145/3190618

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2015). Web-scale training for face identification. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2746–2754. https://doi.org/10.1109/CVPR.2015.7298891

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. Science, 308(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. Psychological Science, 25(7), 1404–1417. https://doi.org/10.1177/0956797614532474

Torralba, A., Fergus, R., & Freeman, W. T. (2008). 80 million tiny images: A large dataset for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11), 1958–1970. https://doi.org/10.1109/TPAMI.2008.128

Wang, D. (2021). Presentation in self-posted facial images can expose sexual orientation: Implications for research and privacy. https://doi.org/10.31234/osf.io/u7vcd

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of Personality and Social Psychology, 114(2), 246–257. https://doi.org/10.1037/pspa0000098

West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. Business & Society, 58(1), 20–41. https://doi.org/10.1177/0007650317718185

White, D., Sutherland, C. A. M., & Burton, A. L. (2017). Choosing face: The curse of self in profile image selection. Cognitive Research: Principles and Implications, 2(1), 23. https://doi.org/10.1186/s41235-017-0058-3

Zhang, Z., Song, Y., & Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. ArXiv:1702.08423 [Cs]. http://arxiv.org/abs/1702.08423

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 Million Image Database for Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

Chapter 3

Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. Journal of Political Economy, 57.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. Journal of Economic Perspectives, 29(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., & von Krogh, G. (2022). We Are All Theorists of Technology Now: A Relational Perspective on Emerging Technology and Organizing. Organization Science, 33(1), 1–18. https://doi.org/10.1287/orsc.2021.1562

Balasubramanian, N., Ye, Y., & Xu, M. (2020). Substituting Human Decision-Making with Machine Learning: Implications for Organizational Learning. Academy of Management Review. https://doi.org/10.5465/amr.2019.0470

Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. WW Norton & Company.

Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. AI Magazine, 26(4), 53–53. https://doi.org/10.1609/aimag.v26i4.1848

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. Strategic Management Journal, 42(1), 30–57. https://doi.org/10.1002/smj.3215

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. Strategic Management Journal, 40(11), 1705–1732. https://doi.org/10.1002/smj.3067

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in Theory Building and Theory Testing: A Five-Decade Study of the Academy of Management Journal. Academy of Management Journal, 50(6), 1281–1303. https://doi.org/10.5465/amj.2007.28165855

Cyert, R. M., Dill, W. R., & March, J. G. (1958). The Role of Expectations in Business Decision Making. Administrative Science Quarterly, 3(3), 307. https://doi.org/10.2307/2390716

Cyert, R. M., & March, J. G. (1963). A Behavioral Theory of The Firm. Wiley-Blackwell.

Cyert, R. M., & March, J. G. (2002). A Summary of Basic Concepts in the Behavioral Theory of the Firm. https://cupdf.com/document/a-summary-of-basic-concepts-in-the-behavioral-theory-of-.html

Davenport, T. H., & College, B. (2016). Only Humans Need Apply: Winners and Losers in the Age of Smart Machines. Harper Business.

De Cremer, D., & De Schutter, L. (2021). How to use algorithmic decision-making to promote inclusiveness in organizations. AI and Ethics. https://doi.org/10.1007/s43681-021-00073-0

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. Strategic Management Journal, 42(12), 2195–2217. https://doi.org/10.1002/smj.3286

Frank, M. R., Sun, L., Cebrian, M., Youn, H., & Rahwan, I. (2018). Small cities face greater impact from automation. Journal of The Royal Society Interface, 15(139), 20170946. https://doi.org/10.1098/rsif.2017.0946

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gavetti, G., Greve, H. R., Levinthal, D. A., & Ocasio, W. (2012). The Behavioral Theory of the Firm: Assessment and Prospects. Academy of Management Annals, 6(1), 1–40. https://doi.org/10.5465/19416520.2012.656841

Gavetti, G., & Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search. Administrative Science Quarterly, 45(1), 25.

Gavetti, G., Levinthal, D., & Ocasio, W. (2007). Perspective—Neo-Carnegie: The Carnegie School's Past, Present, and Reconstructing for the Future. Organization Science, 18(3), 523–536. https://doi.org/10.1287/orsc.1070.0277

Glynn, M. A. (1996). Innovative Genius: A Framework for Relating Individual and Organizational Intelligences to Innovation. Academy of Management Review, 21(4), 1081–1111. https://doi.org/10.5465/amr.1996.9704071864

Gugerty, L. (2006). Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 880–884. https://doi.org/10.1177/154193120605000904

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2020). Ghost in the Machine: On Organizational Theory in the Age of Machine Learning. Academy of Management Review, amr.2019.0247. https://doi.org/10.5465/amr.2019.0247

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2021). The Machine Hums! Addressing Ontological and Normative Concerns Regarding Machine Learning Applications in Organizational Scholarship. Academy of Management Review, amr.2021.0166. https://doi.org/10.5465/amr.2021.0166

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. Organization Science, 33(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Levitt, B., & March, J. G. (1988). Organizational Learning. 23.

Lindebaum, D., & Ashraf, M. (2021). The Ghost In The Machine, Or The Ghost In Organizational Theory? A Complementary View On The Use Of Machine Learning. Academy of Management Review, amr.2021.0036. https://doi.org/10.5465/amr.2021.0036

Lindebaum, D., Vesa, M., & den Hond, F. (2020). Insights From "The Machine Stops" to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations. Academy of Management Review, 45(1), 247–263. https://doi.org/10.5465/amr.2018.0181

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Luan, S., Reb, J., & Gigerenzer, G. (2019). Ecological Rationality: Fast-and-Frugal Heuristics for Managerial Decision Making under Uncertainty. Academy of Management Journal, 62(6), 1735–1759. https://doi.org/10.5465/amj.2018.0172

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. Organization Science, 2(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

March, J. G., & Simon, H. A. (1958). Organizations. John Wiley & Sons. https://papers.ssrn.com/abstract=1496194

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. Psychological Review, 65(3), 151–166. https://doi.org/10.1037/h0048495

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a General Problem-solving Program. Rand Corporation.

Obrad, C., & Circa, C. (2021). Determinants of Work Engagement Among Teachers in the Context of Teleworking. Www.Amfiteatrueconomic.Ro, 23(58), 718. https://doi.org/10.24818/EA/2021/58/718

Puranam, P. (2021). Human–AI collaborative decision-making as an organization design problem. Journal of Organization Design, 10(2), 75–80. https://doi.org/10.1007/s41469-021-00095-2

Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. Academy of Management Review, 46(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Rousseau, D. M. (2011). Reinforcing the Micro/Macro Bridge: Organizational Thinking and Pluralistic Vehicles. Journal of Management, 37(2), 429–442. https://doi.org/10.1177/0149206310372414

Scott, W. R., & Davis, G. F. (2006). Organizations and Organizing: Rational, Natural and Open System Perspectives. Routledge.

Simon, H. A. (1947). Administrative Behavior. Simon and Schuster.
Simon, H. A. (1969). The Sciences of the Artificial (3rd ed.). MIT Press.
Simon, H. A. (1976). From Substantive to Procedural Rationality. In T. J. Kastelein, S. K. Kuipers, W. A. Nijenhuis, & G. R. Wagenaar (Eds.), 25 Years of Economic Theory: Retrospect and prospect (pp. 65–86). Springer US. https://doi.org/10.1007/978-1-4613-4367-7_6
Simon, H. A. (1991). Artificial intelligence: Where has it been, and where is it going? IEEE Transactions on Knowledge and Data Engineering, 3(2), 128–136. https://doi.org/10.1109/69.87993
Simon, H. A., & Newell, A. (1961). GPS, a program that simulates human thought. Rand Corporation.
Stohl, C. (2016). Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age. 15.
Walsh, J. P., & Ungson, G. R. (1991). Organizational memory. The Academy of Management Review, 16(1), 57–91. https://doi.org/10.2307/258607
Wilson, H. J., & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. Harvard Business Review, 11.

Appendix

Table 1

Cronbach's Alpha Reliability of Within-Subject Facial Attributes for Study 1a

Attribute	Women Reliability	95% CI	Men Reliability	95% CI
Neutral	0.670	[0.660, 0.680]	0.681	[0.667, 0.694]
Happiness	0.729	[0.721, 0.737]	0.718	[0.706, 0.730]
Anger	0.141	[0.115, 0.167]	0.354	[0.325, 0.381]
Surprise	0.370	[0.350, 0.389]	0.372	[0.345, 0.399]
Disgust	0.239	[0.216, 0.262]	0.292	[0.261, 0.322]
Sadness	0.228	[0.204, 0.251]	0.257	[0.224, 0.289]
Eyes	0.372	[0.353, 0.391]	0.499	[0.476, 0.520]
Glasses	0.794	[0.787, 0.800]	0.712	[0.699, 0.724]
Smiling	0.766	[0.758, 0.773]	0.726	[0.714, 0.738]
Roll	0.136	[0.110, 0.162]	0.225	[0.191, 0.258]
Yaw	0.327	[0.306, 0.347]	0.327	[0.297, 0.355]
Pitch	0.495	[0.480, 0.510]	0.545	[0.525, 0.565]
Average	0.439	-	0.476	-

Table 2

Sample Size by Cross-Validation Fold, Train-Test Partition and Number of Images Separately for Each Gender for Study 1c

Panel A: Women

Fold	Split	Participants	Images	Participants with:				
				1 Image	2 Images	3 Images	4 Images	⩾5 Images
1	Train	9823	20520	4292	2405	1847	761	518
	Test	517	1080	226	127	97	39	28
2	Train	9823	20520	4292	2405	1847	761	518
	Test	517	1080	226	127	97	39	28
3	Train	9823	20520	4292	2405	1848	759	519
	Test	517	1080	226	127	96	41	27
4	Train	9823	20520	4292	2405	1848	759	519
	Test	517	1080	226	127	96	41	27
5	Train	9823	20520	4292	2405	1848	759	519
	Test	517	1080	226	127	96	41	27
6	Train	9823	20520	4292	2405	1848	759	519
	Test	517	1080	226	127	96	41	27
7	Train	9823	20520	4292	2405	1848	759	519
	Test	517	1080	226	127	96	41	27
8	Train	9823	20520	4292	2405	1848	759	519
	Test	517	1080	226	127	96	41	27
9	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
10	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
11	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
12	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
13	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
14	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
15	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
16	Train	9823	20520	4292	2406	1846	760	519
	Test	517	1080	226	126	98	40	27
17	Train	9823	20520	4292	2406	1846	761	518
	Test	517	1080	226	126	98	39	28
18	Train	9823	20520	4292	2406	1846	761	518
	Test	517	1080	226	126	98	39	28
19	Train	9823	20520	4293	2404	1847	761	518
	Test	517	1080	225	128	97	39	28
20	Train	9823	20520	4293	2404	1847	761	518
	Test	517	1080	225	128	97	39	28
Entire Dataset		10340	21600	4518	2532	1944	800	546

Panel B: Men

Fold	Split	Participants	Images	Participants with:				
				1 Image	2 Images	3 Images	4 Images	⩾5 Images
1	Train	4867	9653	2258	1218	863	336	192
	Test	257	509	120	64	45	18	10
2	Train	4867	9653	2258	1218	863	336	192
	Test	257	509	120	64	45	18	10
3	Train	4868	9654	2259	1218	863	336	192
	Test	256	508	119	64	45	18	10
4	Train	4868	9654	2259	1218	863	336	192
	Test	256	508	119	64	45	18	10
5	Train	4868	9654	2259	1218	863	336	192
	Test	256	508	119	64	45	18	10
6	Train	4868	9654	2259	1218	863	336	192
	Test	256	508	119	64	45	18	10
7	Train	4868	9654	2259	1218	863	336	192
	Test	256	508	119	64	45	18	10
8	Train	4868	9654	2259	1218	863	336	192
	Test	256	508	119	64	45	18	10
9	Train	4868	9654	2259	1219	861	337	192
	Test	256	508	119	63	47	17	10
10	Train	4868	9654	2259	1219	861	337	192
	Test	256	508	119	63	47	17	10
11	Train	4868	9654	2259	1219	861	338	191
	Test	256	508	119	63	47	16	11
12	Train	4868	9654	2259	1219	861	338	191
	Test	256	508	119	63	47	16	11
13	Train	4868	9654	2260	1217	863	336	192
	Test	256	508	118	65	45	18	10
14	Train	4868	9654	2260	1217	863	336	192
	Test	256	508	118	65	45	18	10
15	Train	4868	9654	2260	1217	863	336	192
	Test	256	508	118	65	45	18	10
16	Train	4868	9654	2260	1217	863	336	192
	Test	256	508	118	65	45	18	10
17	Train	4868	9654	2260	1217	863	336	192
	Test	256	508	118	65	45	18	10
18	Train	4868	9654	2260	1217	863	336	192
	Test	256	508	118	65	45	18	10
19	Train	4867	9654	2258	1218	863	336	192
	Test	257	508	120	64	45	18	10
20	Train	4867	9654	2258	1218	863	336	192
	Test	257	508	120	64	45	18	10
Entire Dataset		5124	10162	2378	1282	908	354	202

Table 3

Average AUC Results by Different Degrees of Masking Compared to AUC of Random (Fully Masked Images) for Study 1c

	Women (N = 10,340)				Men (N = 5,124)			
Degree	AUC	95% CI	*p*	*d*	AUC	95% CI	*p*	*d*
Baseline	.702	[.692, .712]	<.001	4.002	.662	[.647, .677]	<.001	2.643
3%	.695	[.685, .705]	<.001	3.856	.660	[.645, .674]	<.001	2.600
6%	.680	[.670, .690]	<.001	3.528	.661	[.646, .675]	<.001	2.616
10%	.665	[.654, .675]	<.001	3.201	.652	[.637, .667]	<.001	2.461
13%	.657	[.646, .667]	<.001	3.035	.644	[.629, .659]	<.001	2.332
16%	.652	[.641, .662]	<.001	2.935	.638	[.623, .653]	<.001	2.225
20%	.647	[.637, .658]	<.001	2.841	.633	[.618, .648]	<.001	2.146
23%	.645	[.635, .656]	<.001	2.800	.628	[.613, .644]	<.001	2.064
26%	.642	[.631, .652]	<.001	2.731	.623	[.608, .638]	<.001	1.975
30%	.639	[.629, .650]	<.001	2.684	.622	[.607, .637]	<.001	1.956
33%	.636	[.625, .646]	<.001	2.606	.619	[.604, .634]	<.001	1.905
36%	.634	[.624, .645]	<.001	2.577	.615	[.600, .630]	<.001	1.842
40%	.631	[.620, .641]	<.001	2.510	.605	[.589, .620]	<.001	1.673
43%	.629	[.619, .640]	<.001	2.477	.606	[.590, .621]	<.001	1.685
46%	.624	[.614, .635]	<.001	2.382	.607	[.591, .622]	<.001	1.701
50%	.620	[.609, .631]	<.001	2.295	.602	[.587, .618]	<.001	1.631
53%	.615	[.604, .626]	<.001	2.192	.595	[.580, .611]	<.001	1.518
56%	.612	[.601, .623]	<.001	2.134	.596	[.581, .612]	<.001	1.533
60%	.605	[.594, .616]	<.001	2.000	.600	[.585, .616]	<.001	1.596
63%	.605	[.594, .616]	<.001	1.996	.593	[.578, .609]	<.001	1.487
66%	.599	[.588, .610]	<.001	1.885	.591	[.576, .607]	<.001	1.449
70%	.592	[.581, .603]	<.001	1.742	.589	[.574, .605]	<.001	1.419
73%	.586	[.575, .597]	<.001	1.632	.589	[.573, .604]	<.001	1.408
76%	.582	[.571, .593]	<.001	1.557	.583	[.568, .599]	<.001	1.324
80%	.575	[.564, .586]	<.001	1.420	.580	[.564, .595]	<.001	1.267
83%	.567	[.556, .578]	<.001	1.260	.579	[.563, .594]	<.001	1.246
86%	.557	[.546, .568]	<.001	1.082	.578	[.563, .594]	<.001	1.241
90%	.554	[.543, .565]	<.001	1.021	.576	[.560, .591]	<.001	1.200
93%	.552	[.541, .563]	<.001	0.975	.575	[.559, .590]	<.001	1.187
96%	.540	[.529, .551]	<.001	0.760	.568	[.552, .583]	<.001	1.072

Note: Average AUC was generated by the following steps. First, averaging the predictions across facial images for the same person. Second, comparing the averaged prediction to the observed value to generate the average AUC.

Table 4

AUC Results Using One Image by Different Degrees of Masking Compared to AUC of Random (Fully Masked Images) for Study 1c

	Women (N = 10,340)				Men (N = 5,124)			
Degree	AUC	95% CI	*p*	*d*	AUC	95% CI	*p*	*d*
Baseline	.687	[.677, .697]	<.001	3.679	.644	[.630, .659]	<.001	2.340
3%	.679	[.669, .689]	<.001	3.514	.644	[.629, .659]	<.001	2.323
6%	.663	[.653, .673]	<.001	3.169	.639	[.624, .654]	<.001	2.244
10%	.648	[.638, .659]	<.001	2.860	.632	[.617, .647]	<.001	2.121
13%	.640	[.629, .650]	<.001	2.693	.628	[.613, .643]	<.001	2.056
16%	.636	[.626, .647]	<.001	2.624	.625	[.610, .640]	<.001	2.008
20%	.632	[.622, .643]	<.001	2.544	.621	[.606, .636]	<.001	1.939
23%	.631	[.620, .642]	<.001	2.513	.614	[.599, .630]	<.001	1.831
26%	.628	[.618, .639]	<.001	2.460	.610	[.595, .626]	<.001	1.765
30%	.626	[.615, .636]	<.001	2.409	.610	[.594, .625]	<.001	1.751
33%	.624	[.613, .635]	<.001	2.376	.605	[.590, .620]	<.001	1.678
36%	.623	[.612, .633]	<.001	2.344	.603	[.588, .618]	<.001	1.643
40%	.620	[.609, .630]	<.001	2.285	.592	[.577, .608]	<.001	1.473
43%	.618	[.607, .628]	<.001	2.246	.592	[.576, .607]	<.001	1.459
46%	.612	[.601, .623]	<.001	2.141	.592	[.577, .608]	<.001	1.472
50%	.608	[.597, .618]	<.001	2.050	.589	[.574, .605]	<.001	1.419
53%	.604	[.593, .614]	<.001	1.972	.582	[.567, .598]	<.001	1.305
56%	.601	[.590, .611]	<.001	1.912	.582	[.566, .597]	<.001	1.302
60%	.594	[.583, .605]	<.001	1.778	.585	[.570, .601]	<.001	1.353
63%	.593	[.582, .604]	<.001	1.758	.578	[.562, .593]	<.001	1.236
66%	.587	[.576, .598]	<.001	1.645	.576	[.560, .592]	<.001	1.206
70%	.580	[.569, .591]	<.001	1.509	.574	[.558, .589]	<.001	1.166
73%	.574	[.564, .585]	<.001	1.408	.573	[.557, .588]	<.001	1.150
76%	.567	[.556, .578]	<.001	1.271	.567	[.552, .583]	<.001	1.069
80%	.563	[.552, .574]	<.001	1.195	.564	[.548, .580]	<.001	1.013
83%	.555	[.544, .566]	<.001	1.032	.564	[.548, .580]	<.001	1.014
86%	.546	[.535, .557]	<.001	0.862	.564	[.549, .580]	<.001	1.017
90%	.545	[.534, .556]	<.001	0.851	.563	[.547, .579]	<.001	0.998
93%	.543	[.532, .555]	<.001	0.818	.561	[.546, .577]	<.001	0.970
96%	.534	[.523, .545]	<.001	0.643	.556	[.540, .572]	<.001	0.884

Table 5

Average Accuracy Results by Different Degrees of Masking for Study 1c

Degree	Women Accuracy	Precision	Recall	F1	Men Accuracy	Precision	Recall	F1
Baseline	64.83%	64.72%	65.18%	64.95%	61.10%	61.21%	60.66%	60.93%
3%	64.49%	64.57%	64.20%	64.38%	61.55%	61.91%	60.07%	60.97%
6%	63.05%	63.32%	62.03%	62.67%	61.94%	62.50%	59.72%	61.08%
10%	61.97%	62.20%	61.04%	61.62%	61.49%	62.35%	58.04%	60.12%
13%	61.40%	61.41%	61.35%	61.38%	60.85%	61.87%	56.56%	59.09%
16%	60.84%	60.92%	60.50%	60.71%	60.13%	61.06%	55.93%	58.38%
20%	60.33%	60.39%	60.02%	60.21%	60.21%	60.91%	56.99%	58.88%
23%	60.09%	60.12%	59.92%	60.02%	59.95%	60.88%	55.70%	58.17%
26%	60.13%	60.18%	59.85%	60.01%	59.93%	61.03%	54.96%	57.84%
30%	60.05%	60.03%	60.14%	60.08%	59.62%	60.41%	55.85%	58.04%
33%	59.73%	59.73%	59.71%	59.72%	58.65%	59.54%	53.98%	56.62%
36%	59.97%	59.87%	60.46%	60.17%	57.77%	58.36%	54.22%	56.21%
40%	59.56%	59.38%	60.56%	59.96%	57.16%	57.55%	54.61%	56.04%
43%	59.26%	59.11%	60.06%	59.58%	57.42%	57.71%	55.50%	56.59%
46%	58.95%	58.75%	60.06%	59.40%	57.90%	58.16%	56.32%	57.23%
50%	58.74%	58.52%	60.08%	59.29%	57.81%	58.01%	56.52%	57.26%
53%	58.11%	57.91%	59.40%	58.65%	56.95%	57.03%	56.36%	56.69%
56%	57.68%	57.40%	59.59%	58.47%	56.81%	56.88%	56.28%	56.58%
60%	57.53%	57.30%	59.17%	58.22%	57.61%	57.67%	57.26%	57.46%
63%	57.33%	57.13%	58.72%	57.92%	56.79%	56.77%	56.99%	56.88%
66%	57.05%	56.88%	58.30%	57.58%	56.42%	56.29%	57.46%	56.87%
70%	56.76%	56.59%	58.05%	57.31%	57.06%	56.80%	59.02%	57.89%
73%	56.19%	56.15%	56.48%	56.32%	56.32%	56.12%	58.00%	57.04%
76%	55.94%	55.89%	56.32%	56.11%	55.58%	55.37%	57.53%	56.43%
80%	55.08%	55.10%	54.89%	54.99%	55.68%	55.42%	58.08%	56.72%
83%	54.57%	54.64%	53.83%	54.23%	55.48%	55.24%	57.77%	56.48%
86%	54.44%	54.64%	52.28%	53.43%	55.21%	55.03%	57.03%	56.01%
90%	54.26%	54.43%	52.32%	53.35%	55.17%	55.01%	56.79%	55.89%
93%	53.86%	53.92%	53.09%	53.50%	55.23%	55.02%	57.34%	56.15%
96%	53.00%	53.09%	51.45%	52.26%	54.25%	54.06%	56.67%	55.34%
Random	50.00%	50.00%	100.00%	66.67%	50.00%	50.00%	100.00%	66.67%

Note: Average accuracy was generated by the following steps. First, averaging the predictions across facial images for the same person. Second, comparing the averaged prediction to the observed value to generate the average accuracy.

Table 6

Accuracy Results for One Image Only by Different Degrees of Masking for Study 1c

Degree	Women Accuracy	Precision	Recall	F1	Men Accuracy	Precision	Recall	F1
0	63.64%	63.39%	64.55%	63.96%	59.37%	59.44%	58.98%	59.21%
1	62.73%	62.68%	62.92%	62.80%	59.84%	60.29%	57.65%	58.94%
2	61.71%	61.77%	61.45%	61.61%	59.31%	59.84%	56.64%	58.19%
3	60.57%	60.68%	60.08%	60.38%	60.03%	60.79%	56.52%	58.58%
4	60.04%	60.02%	60.12%	60.07%	59.15%	59.95%	55.15%	57.45%
5	59.60%	59.53%	59.98%	59.76%	58.63%	59.44%	54.29%	56.75%
6	59.54%	59.51%	59.69%	59.60%	59.17%	59.80%	55.97%	57.82%
7	59.39%	59.43%	59.17%	59.30%	58.39%	59.09%	54.53%	56.72%
8	59.18%	59.21%	59.01%	59.11%	58.80%	59.67%	54.29%	56.86%
9	59.29%	59.25%	59.54%	59.39%	57.85%	58.49%	54.06%	56.19%
10	58.81%	58.76%	59.11%	58.93%	57.20%	58.03%	52.03%	54.87%
11	58.86%	58.78%	59.28%	59.03%	57.01%	57.62%	53.01%	55.21%
12	58.53%	58.36%	59.56%	58.95%	56.05%	56.38%	53.43%	54.87%
13	58.39%	58.24%	59.32%	58.78%	56.36%	56.64%	54.29%	55.44%
14	58.03%	57.83%	59.28%	58.55%	56.26%	56.42%	55.04%	55.72%
15	57.71%	57.46%	59.38%	58.40%	55.84%	56.02%	54.29%	55.14%
16	57.43%	57.22%	58.84%	58.02%	55.58%	55.64%	55.04%	55.34%
17	57.07%	56.82%	58.88%	57.83%	55.39%	55.49%	54.45%	54.96%
18	56.67%	56.42%	58.63%	57.50%	56.30%	56.35%	55.97%	56.16%
19	56.35%	56.15%	58.05%	57.08%	55.25%	55.18%	55.93%	55.55%
20	55.80%	55.68%	56.87%	56.27%	55.00%	54.92%	55.74%	55.33%
21	55.67%	55.49%	57.27%	56.37%	55.41%	55.22%	57.18%	56.18%
22	55.21%	55.17%	55.59%	55.38%	55.21%	55.10%	56.32%	55.70%
23	54.42%	54.35%	55.20%	54.77%	54.02%	53.82%	56.60%	55.18%
24	54.04%	54.01%	54.39%	54.20%	53.96%	53.77%	56.44%	55.08%
25	53.59%	53.66%	52.65%	53.15%	53.98%	53.75%	57.06%	55.36%
26	53.22%	53.37%	51.04%	52.18%	54.47%	54.25%	57.03%	55.60%
27	53.15%	53.30%	50.95%	52.10%	54.31%	54.12%	56.67%	55.37%
28	53.20%	53.25%	52.40%	52.82%	54.63%	54.40%	57.14%	55.74%
29	52.38%	52.46%	50.74%	51.58%	53.69%	53.46%	57.03%	55.18%
30	50.00%	50.00%	100.00%	66.67%	50.00%	50.00%	100.00%	66.67%

Table 7

Statistical Tests of Brightness Differences by Gender and Facial Regions for Study 2a

Panel A: Brightness Difference in Image Background

	Women (N = 10,340)							
	Heterosexual		Lesbian		Significance Test			
	Mean	95% CI	Mean	95% CI	*t*	95% CI	*p*	*d*
Red	0.400	[0.40, 0.40]	0.387	[0.38, 0.39]	-5.328	[-0.02, -0.01]	<.001	0.072
Green	0.352	[0.35, 0.36]	0.342	[0.34, 0.34]	-4.998	[-0.01, -0.01]	<.001	0.068
Blue	0.326	[0.32, 0.33]	0.315	[0.31, 0.32]	-5.137	[-0.01, -0.01]	<.001	0.070
Average	0.359	[0.36, 0.36]	0.348	[0.35, 0.35]	-5.329	[-0.02, -0.01]	<.001	0.073
	Men (N = 5,124)							
	Heterosexual		Gay		Significance Test			
	Mean	95% CI	Mean	95% CI	*t*	95% CI	*p*	*d*
Red	0.423	[0.42, 0.43]	0.443	[0.44, 0.45]	5.532	[0.01, 0.03]	<.001	0.110
Green	0.394	[0.39, 0.40]	0.413	[0.41, 0.42]	5.312	[0.01, 0.03]	<.001	0.105
Blue	0.373	[0.37, 0.38]	0.389	[0.38, 0.39]	4.396	[0.01, 0.02]	<.001	0.087
Average	0.397	[0.39, 0.40]	0.415	[0.41, 0.42]	5.257	[0.01, 0.03]	<.001	0.104

Panel B: Brightness Difference in Facial Region

	Women (N = 10,340)							
	Heterosexual		Lesbian		Significance Test			
	Mean	95% CI	Mean	95% CI	*t*	95% CI	*p*	*d*
Red	0.552	[0.55, 0.55]	0.542	[0.54, 0.54]	-5.664	[-0.01, -0.01]	<.001	0.077
Green	0.418	[0.42, 0.42]	0.412	[0.41, 0.41]	-4.041	[-0.01, -0.00]	<.001	0.055
Blue	0.371	[0.37, 0.37]	0.365	[0.36, 0.37]	-3.453	[-0.01, -0.00]	=.001	0.047
Average	0.447	[0.44, 0.45]	0.440	[0.44, 0.44]	-4.722	[-0.01, -0.00]	<.001	0.064
	Men (N = 5,124)							
	Heterosexual		Gay		Significance Test			
	Mean	95% CI	Mean	95% CI	*t*	95% CI	*p*	*d*
Red	0.522	[0.52, 0.53]	0.549	[0.55, 0.55]	10.896	[0.02, 0.03]	<.001	0.216
Green	0.398	[0.40, 0.40]	0.414	[0.41, 0.42]	7.392	[0.01, 0.02]	<.001	0.147
Blue	0.359	[0.36, 0.36]	0.368	[0.37, 0.37]	4.393	[0.01, 0.01]	<.001	0.087
Average	0.426	[0.42, 0.43]	0.443	[0.44, 0.45]	8.252	[0.01, 0.02]	<.001	0.164

Panel C: Brightness Difference for Entire Image

	Women (N = 10,340)							
	Heterosexual		Lesbian		Significance Test			
	Mean	95% CI	Mean	95% CI	*t*	95% CI	*p*	*d*
Red	0.951	[0.95, 0.96]	0.930	[0.93, 0.93]	-6.568	[-0.03, -0.02]	<.001	0.089
Green	0.771	[0.77, 0.77]	0.754	[0.75, 0.76]	-5.638	[-0.02, -0.01]	<.001	0.077
Blue	0.696	[0.69, 0.70]	0.681	[0.68, 0.68]	-5.325	[-0.02, -0.01]	<.001	0.072
Average	0.806	[0.80, 0.81]	0.788	[0.78, 0.79]	-6.198	[-0.02, -0.01]	<.001	0.084
	Men (N = 5,124)							
	Heterosexual		Gay		Significance Test			
	Mean	95% CI	Mean	95% CI	*t*	95% CI	*p*	*d*
Red	0.946	[0.94, 0.95]	0.992	[0.99, 1.00]	9.701	[0.04, 0.06]	<.001	0.192
Green	0.793	[0.79, 0.80]	0.827	[0.82, 0.83]	7.645	[0.03, 0.04]	<.001	0.152
Blue	0.731	[0.72, 0.74]	0.757	[0.75, 0.76]	5.379	[0.02, 0.03]	<.001	0.107
Average	0.823	[0.82, 0.83]	0.858	[0.85, 0.86]	8.021	[0.03, 0.04]	<.001	0.159

Table 8

Average AUC Results by Different Degrees of Blurring (by Target Width) Compared to AUC of Random (One-pixel Blurred Image of Dataset) for Study 2b

	Women (N = 10,340)				Men (N = 5,124)			
Width	AUC	95% CI	*p*	*d*	AUC	95% CI	*p*	*d*
224	.702	[.692, .712]	<.001	4.010	.660	[.645, .675]	<.001	2.610
112	.703	[.693, .713]	<.001	4.026	.661	[.646, .676]	<.001	2.626
74	.702	[.692, .712]	<.001	4.011	.657	[.643, .672]	<.001	2.562
56	.703	[.693, .713]	<.001	4.020	.662	[.647, .677]	<.001	2.641
44	.702	[.692, .712]	<.001	4.002	.668	[.653, .683]	<.001	2.745
37	.697	[.687, .707]	<.001	3.889	.662	[.647, .676]	<.001	2.633
32	.691	[.681, .701]	<.001	3.770	.660	[.645, .674]	<.001	2.596
28	.689	[.679, .699]	<.001	3.722	.657	[.642, .672]	<.001	2.553
24	.680	[.670, .691]	<.001	3.535	.645	[.630, .660]	<.001	2.350
22	.670	[.660, .680]	<.001	3.311	.642	[.627, .658]	<.001	2.302
20	.656	[.646, .667]	<.001	3.032	.636	[.620, .651]	<.001	2.184
19	.656	[.646, .667]	<.001	3.022	.631	[.616, .646]	<.001	2.104
18	.646	[.636, .657]	<.001	2.821	.618	[.603, .633]	<.001	1.888
17	.642	[.631, .652]	<.001	2.731	.615	[.599, .630]	<.001	1.835
16	.623	[.612, .634]	<.001	2.356	.594	[.579, .610]	<.001	1.501
15	.618	[.607, .628]	<.001	2.249	.601	[.585, .616]	<.001	1.607
14	.614	[.603, .624]	<.001	2.168	.585	[.570, .601]	<.001	1.352
13	.607	[.596, .618]	<.001	2.044	.589	[.573, .604]	<.001	1.413
12	.599	[.588, .610]	<.001	1.879	.584	[.569, .600]	<.001	1.337
11	.596	[.585, .607]	<.001	1.818	.581	[.565, .596]	<.001	1.281
10	.588	[.577, .599]	<.001	1.666	.585	[.569, .600]	<.001	1.347
9	.590	[.579, .600]	<.001	1.698	.580	[.565, .596]	<.001	1.272
8	.583	[.572, .594]	<.001	1.568	.586	[.571, .602]	<.001	1.374
7	.581	[.570, .592]	<.001	1.530	.589	[.573, .604]	<.001	1.414
6	.572	[.561, .583]	<.001	1.355	.593	[.578, .609]	<.001	1.484
5	.575	[.564, .586]	<.001	1.421	.578	[.563, .594]	<.001	1.241
4	.558	[.547, .569]	<.001	1.090	.571	[.555, .586]	<.001	1.122
3	.559	[.548, .570]	<.001	1.106	.566	[.550, .581]	<.001	1.038
2	.533	[.522, .544]	<.001	0.628	.568	[.552, .584]	<.001	1.078
1	.529	[.518, .540]	<.001	0.546	.563	[.548, .579]	<.001	1.000

Note: Average AUC was generated by the following steps. First, averaging the predictions across facial images for the same person. Second, comparing the averaged prediction to the observed value to generate the average AUC.

Table 9

AUC Results Using One Image by Different Degrees of Blurring (by Target Width) Compared to AUC of Random (One-pixel Blurred Image of Dataset) for Study 2b

	Women (N = 10,340)				Men (N = 5,124)			
Width	AUC	95% CI	*p*	*d*	AUC	95% CI	*p*	*d*
224	.687	[.677, .697]	<.001	3.681	.643	[.628, .658]	<.001	2.314
112	.688	[.678, .698]	<.001	3.696	.643	[.628, .658]	<.001	2.315
74	.687	[.677, .697]	<.001	3.681	.642	[.627, .657]	<.001	2.290
56	.687	[.677, .698]	<.001	3.691	.647	[.632, .662]	<.001	2.376
44	.686	[.676, .696]	<.001	3.652	.646	[.631, .661]	<.001	2.369
37	.680	[.670, .690]	<.001	3.528	.640	[.625, .655]	<.001	2.261
32	.673	[.663, .683]	<.001	3.380	.638	[.623, .653]	<.001	2.228
28	.670	[.660, .681]	<.001	3.325	.638	[.623, .653]	<.001	2.224
24	.660	[.650, .670]	<.001	3.109	.623	[.608, .638]	<.001	1.976
22	.650	[.640, .661]	<.001	2.904	.619	[.603, .634]	<.001	1.901
20	.637	[.627, .648]	<.001	2.643	.617	[.602, .633]	<.001	1.881
19	.634	[.623, .644]	<.001	2.569	.610	[.595, .625]	<.001	1.757
18	.630	[.619, .641]	<.001	2.490	.601	[.586, .616]	<.001	1.613
17	.620	[.609, .631]	<.001	2.298	.598	[.582, .613]	<.001	1.559
16	.606	[.595, .617]	<.001	2.023	.581	[.566, .597]	<.001	1.289
15	.598	[.587, .609]	<.001	1.869	.586	[.570, .602]	<.001	1.367
14	.594	[.583, .605]	<.001	1.789	.571	[.555, .586]	<.001	1.121
13	.594	[.583, .605]	<.001	1.779	.575	[.559, .590]	<.001	1.182
12	.585	[.574, .596]	<.001	1.610	.573	[.558, .589]	<.001	1.159
11	.582	[.571, .593]	<.001	1.556	.570	[.554, .585]	<.001	1.104
10	.575	[.564, .585]	<.001	1.408	.571	[.555, .586]	<.001	1.123
9	.576	[.565, .587]	<.001	1.441	.566	[.550, .581]	<.001	1.041
8	.569	[.558, .580]	<.001	1.300	.570	[.554, .585]	<.001	1.105
7	.567	[.556, .578]	<.001	1.272	.574	[.559, .590]	<.001	1.177
6	.562	[.551, .573]	<.001	1.177	.581	[.566, .597]	<.001	1.291
5	.565	[.554, .576]	<.001	1.228	.564	[.549, .580]	<.001	1.021
4	.550	[.539, .561]	<.001	0.948	.559	[.544, .575]	<.001	0.936
3	.546	[.535, .557]	<.001	0.872	.552	[.536, .568]	<.001	0.819
2	.527	[.516, .538]	<.001	0.509	.548	[.532, .564]	<.001	0.760
1	.523	[.512, .534]	<.001	0.437	.548	[.532, .564]	<.001	0.758

Table 10

Average Accuracy Results by Different Degrees of Blurring (by Target Width) for Study 2b

Width	Women Accuracy	Precision	Recall	F1	Men Accuracy	Precision	Recall	F1
224	64.76%	64.63%	65.20%	64.91%	61.32%	61.58%	60.19%	60.88%
112	64.93%	64.79%	65.40%	65.09%	61.65%	61.94%	60.46%	61.19%
74	64.89%	64.71%	65.53%	65.12%	61.28%	61.50%	60.30%	60.90%
56	64.92%	64.69%	65.73%	65.20%	61.44%	61.59%	60.77%	61.18%
44	64.76%	64.59%	65.34%	64.96%	62.24%	62.46%	61.36%	61.90%
37	64.11%	63.76%	65.40%	64.57%	61.59%	61.56%	61.75%	61.65%
32	64.52%	64.16%	65.76%	64.95%	62.06%	62.37%	60.81%	61.58%
28	63.97%	63.58%	65.44%	64.49%	61.44%	61.84%	59.72%	60.76%
24	63.42%	63.04%	64.89%	63.95%	60.40%	60.85%	58.35%	59.57%
22	62.66%	62.23%	64.43%	63.31%	60.85%	61.53%	57.92%	59.67%
20	61.42%	61.01%	63.31%	62.14%	60.44%	61.19%	57.10%	59.08%
19	62.26%	61.60%	65.11%	63.31%	59.89%	60.57%	56.71%	58.58%
18	61.33%	60.75%	64.06%	62.36%	58.24%	58.54%	56.44%	57.47%
17	60.38%	59.86%	63.02%	61.40%	58.70%	59.18%	56.13%	57.61%
16	59.31%	58.56%	63.69%	61.02%	56.79%	56.79%	56.83%	56.81%
15	58.17%	57.71%	61.18%	59.39%	56.91%	57.21%	54.84%	56.00%
14	58.14%	57.68%	61.16%	59.37%	56.64%	56.69%	56.25%	56.47%
13	57.55%	57.17%	60.19%	58.65%	56.44%	56.47%	56.21%	56.34%
12	57.43%	57.09%	59.79%	58.41%	56.03%	55.92%	56.99%	56.45%
11	57.01%	56.77%	58.82%	57.78%	56.28%	56.20%	56.95%	56.57%
10	56.61%	56.33%	58.76%	57.52%	56.67%	56.42%	58.63%	57.50%
9	56.23%	55.98%	58.28%	57.11%	56.19%	56.05%	57.30%	56.67%
8	56.32%	56.22%	57.08%	56.65%	56.23%	56.12%	57.10%	56.61%
7	55.84%	55.69%	57.16%	56.41%	56.83%	56.81%	56.99%	56.90%
6	55.05%	54.91%	56.50%	55.69%	56.93%	56.85%	57.53%	57.19%
5	55.55%	55.58%	55.32%	55.45%	55.89%	55.73%	57.30%	56.51%
4	54.04%	54.02%	54.29%	54.16%	54.98%	54.93%	55.46%	55.20%
3	54.35%	54.15%	56.77%	55.43%	54.88%	54.86%	55.07%	54.97%
2	52.79%	52.76%	53.29%	53.02%	55.13%	55.10%	55.46%	55.28%
1	52.34%	52.36%	51.99%	52.17%	54.18%	54.16%	54.33%	54.25%
Random	50.00%	50.00%	100.00%	66.67%	50.00%	50.00%	100.00%	66.67%

Note: Average accuracy was generated by the following steps. First, averaging the predictions across facial images for the same person. Second, comparing the averaged prediction to the observed value to generate the average accuracy.

Table 11

Accuracy Results for One Image Only by Different Degrees of Blurring (by Target Width) for Study 2b

Width	Women Accuracy	Precision	Recall	F1	Men Accuracy	Precision	Recall	F1
224	63.83%	63.59%	64.70%	64.14%	59.70%	59.89%	58.74%	59.31%
112	63.38%	63.10%	64.41%	63.75%	60.19%	60.47%	58.82%	59.64%
74	63.63%	63.29%	64.87%	64.07%	60.25%	60.50%	59.02%	59.75%
56	63.40%	63.09%	64.60%	63.84%	60.62%	60.85%	59.56%	60.20%
44	63.41%	63.09%	64.64%	63.86%	60.34%	60.56%	59.33%	59.94%
37	63.00%	62.53%	64.85%	63.67%	59.78%	59.91%	59.13%	59.52%
32	62.75%	62.38%	64.24%	63.29%	60.34%	60.77%	58.35%	59.54%
28	62.36%	61.91%	64.24%	63.05%	59.62%	59.99%	57.77%	58.86%
24	61.76%	61.28%	63.89%	62.56%	58.43%	58.83%	56.17%	57.47%
22	60.84%	60.22%	63.87%	61.99%	58.47%	59.12%	54.92%	56.94%
20	59.36%	58.94%	61.72%	60.30%	58.10%	58.76%	54.33%	56.46%
19	59.95%	59.37%	63.08%	61.16%	57.75%	58.21%	54.92%	56.52%
18	59.84%	59.27%	62.86%	61.02%	56.71%	56.94%	55.07%	55.99%
17	58.32%	57.90%	60.97%	59.39%	56.83%	57.05%	55.27%	56.15%
16	57.91%	57.32%	61.97%	59.55%	55.99%	56.06%	55.46%	55.76%
15	56.91%	56.54%	59.73%	58.09%	55.89%	56.13%	53.98%	55.03%
14	56.72%	56.35%	59.63%	57.95%	55.21%	55.22%	55.15%	55.18%
13	56.46%	56.14%	59.03%	57.55%	55.00%	54.99%	55.04%	55.01%
12	56.25%	56.02%	58.14%	57.06%	55.09%	54.97%	56.36%	55.66%
11	55.93%	55.75%	57.52%	56.62%	55.37%	55.23%	56.64%	55.93%
10	55.34%	55.17%	57.00%	56.07%	55.27%	55.05%	57.42%	56.21%
9	55.08%	54.90%	56.87%	55.87%	55.13%	54.92%	57.30%	56.08%
8	55.28%	55.16%	56.48%	55.81%	54.78%	54.68%	55.89%	55.28%
7	54.85%	54.80%	55.44%	55.12%	55.52%	55.37%	56.99%	56.16%
6	54.36%	54.24%	55.82%	55.02%	56.67%	56.58%	57.38%	56.98%
5	54.65%	54.69%	54.27%	54.48%	55.31%	55.16%	56.71%	55.93%
4	53.33%	53.33%	53.25%	53.29%	54.84%	54.77%	55.54%	55.16%
3	53.17%	53.02%	55.69%	54.32%	53.98%	53.96%	54.29%	54.12%
2	51.81%	51.78%	52.65%	52.21%	53.01%	52.98%	53.43%	53.21%
1	51.94%	51.96%	51.47%	51.72%	53.10%	53.11%	53.01%	53.06%
Random	50.00%	50.00%	100.00%	66.67%	50.00%	50.00%	100.00%	66.67%

Figure 1

An Example Rendering of Image Augmentations Applied in Studies 1c and 2b

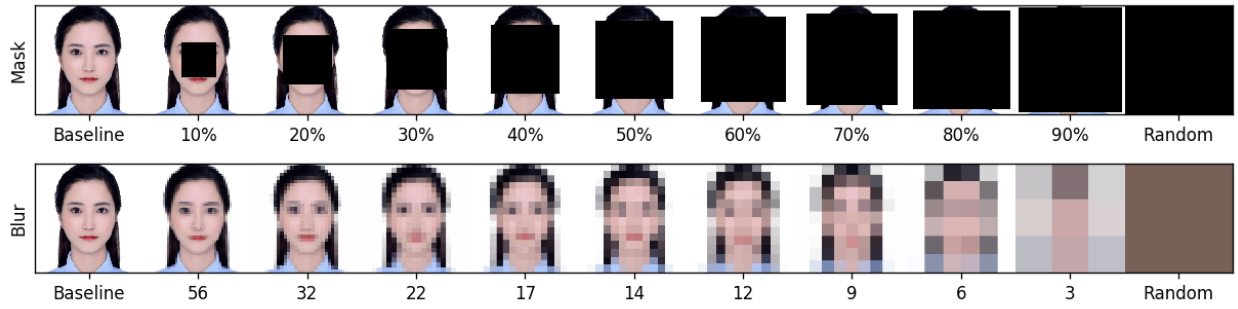


Figure 2

A Plot of the Target Width Used to Downsize Images for Study 2b