

AGE ESTIMATION OF STRANGERS IN PUBLIC SETTINGS: A PROOF-OF-CONCEPT FIELD STUDY

Lasse Suonperä Liebst^{*1,2} , Lasse Baggesen³ , Kasper Lykke Dausel⁴ , Marie
Rosenkrantz Lindegaard⁵ 

¹Department of Sociology, University of Copenhagen, Denmark

²NSCR: The Netherlands Institute for the Study of Crime and Law Enforcement, Netherlands

³ROCKWOOL Foundation Interventions Unit, Denmark

⁴Department for Pediatrics and Adolescent Medicine, Rigshospitalet, Denmark

⁵Department of Sociology, Vrije Universiteit Amsterdam, Netherlands

^{*}ls@soc.ku.dk

ABSTRACT

Laboratory research has shown that age can be estimated with mean absolute errors of approximately five years for faces and ten years for voices. However, it is less clear how precise such age estimates are in real-world settings, where observers often judge from a distance with less controlled and more variable access to facial and vocal cues. This proof-of-concept study assessed the feasibility of examining age estimation under ecologically valid conditions through naturalistic observation in public settings. Two observers estimated the ages of 565 strangers in public from a distance and subsequently approached each individual to record their self-reported age. The findings demonstrate that age estimation can be feasibly conducted in a systematic way through naturalistic observation, with preliminary evidence suggesting that the resulting mean absolute error was comparable to—and possibly lower than—errors reported in laboratory studies. While tentative, these results highlight the potential of studying age estimation in public contexts and support further observational research with larger and more diverse samples.

Keywords: age estimation, naturalistic observation, ecological validity, public settings

Theoretical frameworks across the social and behavioral sciences agree that when we meet another person for the first time, we quickly form an initial characterization of them based on salient behavioral and bodily features—including their age (Fiske, 2018; Turner, 2010). This social categorization shapes how we perceive strangers and how we interact with them. One example, pronounced in some cultural contexts, is that politeness norms require people to adjust their manners and forms of address according to the age difference between interlocutors (e.g., Yoon, 2004). Another example of age-based categorization is in the sale of alcohol and cigarettes, where sales staff are legally required to make rapid assessments of a customer's age to determine whether a purchase is permitted (Kask & Tael-Oeren, 2017).

In line with these examples, a substantial body of evidence suggests that humans can often estimate the chronological age of strangers within a relatively small margin of error (for comprehensive reviews, see Moyse, 2014; Rhodes, 2009). This is typically examined by instructing participants to age-estimate images of faces or recordings of voices. Both faces and voices provide physiological cues of age, such as graying hair, facial wrinkling (Lowry, 2020), and increased instability in voice pitch and loudness (Dehqan et al., 2012). In general, age is more accurately estimated from faces than voices—the mean absolute error between estimated and actual age is typically around five years for faces and 10 years for voices (Moyse, 2014). Further, various group characteristics of observers and subjects, such as their age, ethnicity, and gender (Voegeli et al., 2021), can impact the accuracy of age estimates. For example, studies have identified an own-age bias, whereby people estimate the ages of individuals from their own age group more accurately than those from other age groups (Rhodes & Anastasi, 2012).

These findings on human age estimation accuracy have paralleled and informed developments in computer vision, where researchers have sought to automate estimations. Studies within this line of research have shown that computer-based age estimation from facial images often has accuracy comparable to that of humans (Alarifi et al., 2019). However, there are notable variations between studies, with some showing that computers can outperform humans (Han et al., 2013), and others finding that computers are less accurate and more biased than humans (Ganel et al., 2022). Additionally, computer vision research has estimated age from gait (i.e., manner of walking) and speech features, but with lower accuracy than face-based estimates (Punyani et al., 2018).

Most existing evidence on human and computer-based age estimation comes from controlled laboratory studies, raising questions about the ecological validity of these findings (Shamay-Tsoory & Mendelsohn, 2019). For example, human error in age estimates observed in laboratory settings may partly reflect serial dependency biases, whereby judgments are systematically influenced by the age of the immediately preceding face (Clifford et al., 2018). Further, the accuracy of estimating high-quality facial images and voice recordings in laboratory settings may be inflated compared to estimates made in real-world settings, where observers have less controlled and more variable access to facial and vocal cues. This issue is indicated in the computer vision field, where research has shown that when faces are observed at a distance—a typical condition in naturalistic public settings—the accuracy of face-based age estimation decreases, since facial features become unreliable (Yuan et al., 2018).

In theory, age estimation from a distance in real-life public settings may be more precise, as such contexts provide a variety of alternative cues that can compensate for limited access to facial and vocal information. This expectation aligns with the Brunswikian Lens Model, which holds that when particularly informative cues—such as detailed facial features—are only partially available, observers can draw on and integrate multiple alternative cues to arrive at an accurate judgment (Brunswik, 1952; Wolf, 1999). This is an adaptive feature of human perception that enables humans to survive and act effectively in the face of environmental uncertainty and incomplete information. Estimating chronological age has been highlighted as a core example of the Brunswikian Lens Model, as age estimation plausibly involves the

combination of cues—facial, vocal, bodily, postural, contextual, and clothing-related—that together support accurate estimation (Hastie & Dawes, 2009). For example, contextual indicators such as the tendency for people in their twenties to gather in certain urban areas, the presence of age-typical activities like walking with a stroller, or generational fashions may complement physical cues and substitute for restricted access to facial and vocal detail.

Overall, there are compelling methodological and theoretical reasons to examine the ecological validity of the predominantly laboratory-based age estimation literature. To date, such an examination largely remains to be undertaken. This body of work also shares the broader limitations of social and behavioral research, which has been extensively criticized for its often low ecological validity (Vazire et al., 2022).

The purpose of this paper is to explore the feasibility and relevance of adopting a more ecologically valid approach to studying age estimation in real-world contexts. To this end, we present a proof-of-concept study that develops and tests an on-site observational approach to examining age estimation from a distance in public settings. This design provides the potential to compare laboratory-based and observational age estimation, thereby helping to validate laboratory findings and enhance the ecological validity of the age-estimation literature.

METHOD

Data and sampling

The empirical material consisted of individuals observed in public spaces in Copenhagen, Denmark, during the fall of 2019 (codebook, questionnaire, data, scripts, and full outputs are available at <https://osf.io/h3u74>). To ensure variation, observations were conducted at 52 different locations across the city, including parks, the central station, shopping malls, and recreational nightlife areas. Data collection took place on both weekdays and weekends, spanning mornings, afternoons, and late evenings.

The dataset was originally collected with the additional aim of examining whether social relations could be inferred from interpersonal behavior; results of that analysis are reported elsewhere (Liebst et al., 2023). Accordingly, sampling was based on randomly selected pairs of individuals co-present in the same location. While some of these pairs were socially affiliated, this relational aspect is not relevant here, as the present study focuses solely on age estimation at the individual level¹. In total, we observed 565 individuals across 36 observation days. We spent more than 250 hours in the field, of which approximately 24 hours consisted of active observation of individuals, with the remainder spent on logistics, breaks, identifying participants, and post-observation procedures.

The use of naturalistic observations in public places is consistent with the ethical guidelines of the American Sociological Association (2018, Section 11.1b). The study was approved by the Ethics Review Board at the Department of Sociology, University of Copenhagen.

Coding procedure

For each selected person, covert observations were designed to last 1–5 minutes, mimicking the brief periods during which strangers are typically observed in everyday public settings. This corresponds to a focal sampling strategy, where individuals are observed continuously for a defined period (Altmann, 1974). During these observations, data were recorded using a codebook containing detailed behavioral definitions and observer instructions, in line with methodological guidelines from the ethological tradition of naturalistic observation (Altmann, 1974; Jones et al., 2018). The codebook was developed through weeks of qualitative field observations, which allowed us to familiarize ourselves with

the observation settings, develop behavioral codes and procedures, and iteratively refine the codebook design.

All observations reported in the present paper were conducted by the two junior researchers—both co-authors—who had developed and refined the codebook, thereby acquiring extensive training with it. This means that observer selection constituted a convenience sample, as the two observers were chosen for their role as researchers rather than through random sampling. The observers were male university students in their twenties residing in Copenhagen.

After developing the codebook, we assessed its interrater reliability by double-coding 86 individuals, which is well above the commonly recommended minimum of 30 for agreement testing (McHugh, 2012). To ensure comparability, the two observers coordinated on-site to identify the same individuals, who were then observed independently. These individuals were sampled from 15 urban settings and varied across weekdays, weekends, and times of day. The results of this reliability test are reported in the Results section. Outside this interrater reliability subset, all observations were conducted by one of the two observers.

The codebook included a range of behavioral measures; however, for the present study, the key measure was each observers' best-guess estimate of the individual's age, based on all available situational information. By employing this open-ended age definition, we allowed for the natural variations in the sensory cues available when observing strangers in public (Goffman, 1971). In practice, this meant that in some cases observers could view the individual's facial appearance, though always from a distance that typically limited access to finer facial details. The observation distance was roughly 20 to 30 meters. Observers could not hear what the person was saying, although in some cases fragmented sound cues were audible. However, they could always observe various nonverbal aspects, such as the person's clothing, behavior, and the setting in which the observation occurred—all of which may carry cues about age and could be used to estimate the person's age (Hastie & Dawes, 2009).

After each observation had concluded, the observed person was approached on site and asked a range of questions, including their actual chronological age. At this point, we briefly introduced the purpose of the project, with the additional aim of obtaining informed consent for inclusion in the study. Although we did not systematically record a response rate, our field experience indicated that the vast majority of individuals approached agreed to participate.

RESULTS

The median age of the sampled individuals was 30 years (range = 18–93). A nonparametric sign test ($p < .001$) suggested that this median age was lower than the adult population median of 37 years in Copenhagen Municipality, where all sampling was conducted. The average observation time was 3.6 minutes ($SD = 1.4$, range = 0.67–5.0).

The interrater reliability of the observers in judging age yielded a Krippendorff's (2004) alpha of .91 (95% CI [.88, .95]), exceeding the .75 threshold typically regarded as excellent (Fleiss, 1981). We then examined whether this interrater reliability was accompanied by high accuracy in estimating the actual age. This was supported by an alpha of .93 (95% CI [.92, .95]) for the association between the estimated and actual age. However, the magnitude of this latter alpha score was sensitive to the weighting scheme used to define how errors are penalized (Klein, 2018). In this calculation, we applied a quadratic weight, which penalizes minor disagreements (e.g., 20 vs. 21) less than larger ones (e.g., 20 vs. 22). By contrast, when a linear weight was applied—penalizing disagreements in direct proportion to their size—the alpha decreased to .77 (95% CI [.75, .79]), only just above the cited threshold of .75 for excellent agreement.

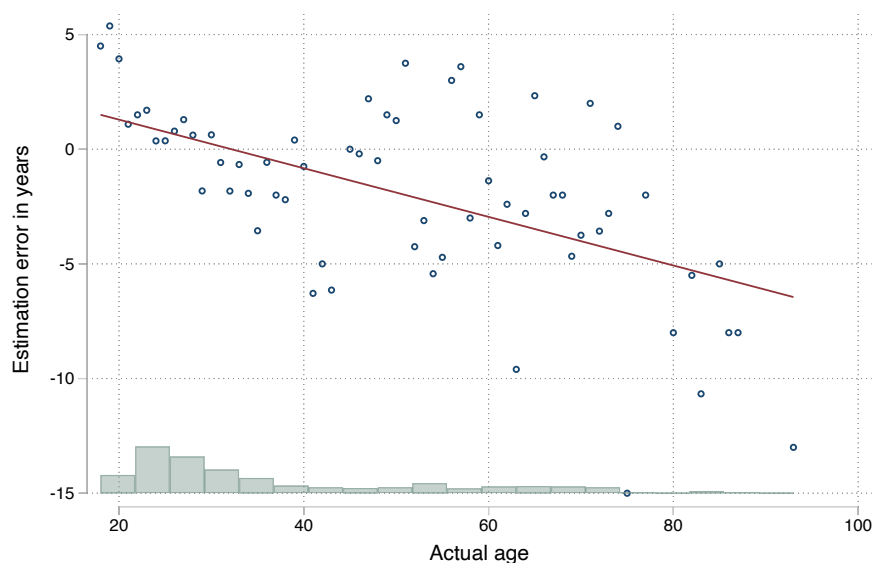
Another and more commonly used measure of age estimation accuracy is the mean absolute error (MAE) between the estimated and actual age (Agbo-Ajala et al., 2022). In our data, the mean absolute error was 4.0 years ($SD = 4.1$), with a 95% confidence interval ranging from 3.6 to 4.3. For comparison, typical mean absolute errors reported in the laboratory-based literature—around 5 years for face-based estimates and around 10 years for voice-based estimates—fall above the upper bound of this confidence interval. This indicates that age estimation may be more accurate under naturalistic observation conditions compared to laboratory-based settings.

Because mean absolute error is sensitive to outliers (Wang et al., 2017), we examined its distribution. In our data, 3.4% of cases were mild outliers and 1.1% were severe outliers, defined as observations beyond Tukey's inner and outer fences, respectively (Dawson, 2011). These correspond to instances where raters made very large errors in estimating age, with mild outliers deviating by more than 11 years and severe outliers by more than 17 years. As an outlier-robust alternative (Wang et al., 2017), the median absolute error (MedAE) was 3 years (bootstrapped 95% CI [2.9, 3.1]).

Next, the absolute error in estimations was found to increase with the subjects' actual age, as shown by a statistically significant regression coefficient, $B = 0.06$, 95% CI [0.04, 0.07], $p < .001$. This indicates that the accuracy of age estimation declines as individuals get older. All regression models used robust regression to account for outliers in age estimation errors (Mehmetoglu & Jakobsen, 2017). In terms of the magnitude of the positive effect of actual age on the absolute error in years, the error was predicted to be 2.3 for 20-year-old subjects and 4.5 for 60-year-old subjects. Further, Figure 1 illustrates the pattern underlying the positive regression slope: observers systematically overestimated ages when judging younger individuals and underestimated ages when judging older individuals.

Figure 1.

Scatterplot of estimation error in years by actual age.



Note. The figure was constructed with a binned scatterplot (Starr & Goldfarb, 2020), which helps visualize associations without distortion from outliers. Each point represents the average estimation error in years for a given actual age value. The histogram along the x-axis shows the distribution of actual age.

We also assessed whether the absolute error between the actual and estimated age was associated with the observation time in minutes, as a longer observation time could potentially provide the observer with more information to make an accurate estimate. However, this association was not statistically significant, $B = -0.12$, 95% CI $[-0.30, 0.05]$, $p = .151$. Moreover, we tested whether absolute error was associated with how many days into the sampling period data were collected, as observers might become more accurate with practice over time. This test was also not statistically significant, $B = -0.01$, 95% CI $[-0.06, 0.04]$, $p = .732$.

DISCUSSION

The purpose of this proof-of-concept study was to evaluate the feasibility of examining age estimations through naturalistic public observations, thereby providing a basis for assessing the ecological validity of findings from predominantly laboratory-based research. Overall, our approach appeared feasible: the codebook could be systematically applied in field settings, as demonstrated by high interrater reliability, and the questionnaire instruments were efficiently employed.

What further underscores the relevance of our observational method for studying age estimation is that the approach yields results that are both meaningful and complementary to the laboratory-based literature. Our age estimates showed a mean absolute error comparable to—and possibly lower than—scores reported in face- and voice-based laboratory studies. This provides tentative evidence that laboratory-based results appear generalizable to real-world settings. Moreover, consistent with previous laboratory findings, we also found that the accuracy of age estimates was higher for younger than for older individuals (Moyse, 2014), and that younger persons were overestimated while older persons were underestimated (Clifford et al., 2018). Replicating these detailed patterns from the existing literature demonstrates that our approach can generate valid and interpretable findings.

While our design proved feasible and capable of yielding informative results, its implementation also revealed a practical challenge: the substantial labor required to carry it out at scale. In our case, more than 250 hours were spent in the field, yet the effective observation time amounted to only about 24 hours. This imbalance—long recognized in the observational methods literature (Altmann, 1974)—underscores the considerable resources required for field-based observation. This time imbalance may incentivize researchers to favor laboratory-based approaches, not because they are more appropriate but because they allow data to be obtained more quickly (see Cialdini, 2009).

Implications

The current findings have several implications. First, the fact that observers achieved age estimation accuracy comparable to laboratory research—despite limited access to detailed facial and vocal cues—suggests a need to reconsider how age estimation functions in naturalistic settings. The Brunswikian Lens Model provides a plausible framework: when access to one type of cue is restricted, observers may compensate by drawing on multiple alternative cues (Brunswik, 1952; Wolf, 1999). While facial appearance is a particularly informative age cue, our findings indicate that comparable accuracy can also be achieved by attending to the broader range of cues available in public encounters, encompassing the person's "setting, appearance, and manner" (Goffman, 1956, p. 16). In our study, such cues may have included age-typical styles of dress (Twigg, 2015) and public behaviors (Marcum, 2013), as well as local knowledge of where different age groups tend to congregate (Santani et al., 2016). For example, a substantial proportion of our data was collected in recreational nightlife

areas known to attract younger people—knowledge that was readily available to our observers as residents of the city.

At the same time, it should be acknowledged that our design does not reveal which specific cues—or which combinations of cues—informed the judgments, since observers were simply instructed to use any information available in the situation. Future research should therefore disentangle the relative importance and interplay of different cue types in real-life public settings.

Second, our results are encouraging for observational research involving age estimation in naturalistic contexts where direct age measurement is not possible. This includes studies in small-scale societies where individuals may not track their age in calendar years (Diekmann et al., 2017), and observational studies relying solely on public security camera footage (Liebst et al., 2022). In such contexts, a key methodological issue arises: researchers should not rely solely on interrater reliability tests of age estimates, as is commonly done. High interrater reliability indicates consistency across observers, but it does not guarantee validity or accuracy, as observers may simply be systematically wrong in their evaluations (Menon et al., 2025). We therefore recommend that observational studies capturing age data combine interrater reliability assessments with accuracy validation tests, using methodological designs similar to the one presented here.

Third, our preliminary results may have practical implications for both professionals and systems that rely on accurate age estimation—for example, police officers identifying suspects (Anda et al., 2018), online age verification mechanisms (Jarvie & Renaud, 2024), and sales staff handling age-restricted products such as alcohol and tobacco (Kask & Tael-Oeren, 2017). For sales contexts, it is noteworthy that age estimates are more accurate for younger individuals, precisely where evaluations of legal purchasing age are most consequential. Although we found a slight tendency to overestimate the ages of younger individuals (see also Moyse, 2014), the widespread violations of age restrictions are thus unlikely to be explained by misjudgments of age alone. They should also be understood in relation to sociological factors, such as a lack of motivation to enforce restrictions despite recognizing underage customers (Gosselt et al., 2012).

Limitations

The main limitation of this study—and the reason it is a proof-of-concept rather than a full-scale investigation—is the reliance on a very small, non-randomly selected sample of only two raters. While they produced informative estimates with an error comparable to that reported in laboratory research, the nature of the sample severely limits the generalizability of the findings. Furthermore, because both raters shared similar demographics, training, and experience, we were unable to assess potential between-rater heterogeneities in age-estimation accuracy. For example, it is plausible that untrained members of the public would provide less accurate judgments (Sörqvist & Eriksson, 2007). While we found no evidence that accuracy improved over the sampling period, both raters had already undergone several weeks of codebook development and training before data collection began, setting them apart from typical laypersons. In addition, both raters were in their twenties, which may have inflated the observed association between actual age and absolute error, consistent with an own-age bias whereby accuracy is higher when judging individuals closer to one's own age (Rhodes & Anastasi, 2012). Future large-scale studies should therefore recruit a broader and more diverse pool of raters to strengthen generalizability and enable systematic examination of between-rater heterogeneities.

Another related limitation concerns the generalizability of our findings across contexts and the demographics of those observed. Because younger individuals' ages are generally estimated with greater accuracy, the lower median age of our sample compared to the general adult population of Copenhagen could have deflated the reported mean absolute error. Additionally,

we did not consider how age estimation accuracy might be influenced by properties of the setting or of the individuals being judged, including the wearing of sunglasses in summer (Thorley, 2021), face masks during the COVID-19 pandemic (Thorley et al., 2022), differences in ethnic background (Voegeli et al., 2021), or the presence of other faces in the vicinity (Awad et al., 2020). Finally, it would be valuable to vary the distance between observers and subjects. In the present study, observations were restricted to distant viewing without direct face-to-face interaction, which limited access to fine-grained facial and vocal cues. By contrast, observers who interact directly with strangers in public may gain richer and more dynamic information than is typically available in laboratory settings. Future research should explore age estimation across a broader range of contexts and interactional conditions, and should ideally include culturally and nationally diverse settings (Vazire et al., 2022).

Conclusion

This proof-of-concept study demonstrated that age estimations from a distance can be systematically and reliably conducted through naturalistic observations in real-world public settings. The preliminary finding that such estimates achieve an accuracy comparable to or exceeding laboratory results provides tentative support for the ecological validity of laboratory-based research. Building on and adapting the methodological approach developed here, future studies should test whether these patterns replicate in larger and more diverse samples.

NOTE

¹One concern is that sampling two individuals at a time could violate the statistical assumption of independent observations. For example, if two people were observed together in the same group, the estimate for one person might influence the estimate for the other because their shared group membership could lead the observer to assume they are similar in age. To address this, we reanalyzed the data using only one individual from each selected pair. This reanalysis did not meaningfully alter any of the reported results.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

DECLARATION

We declare that all work herein is original and our own. All authors have reviewed and approved the manuscript for submission, and it is not being considered for submission elsewhere whilst it is under consideration for *Human Ethology*.

TRANSPARENCY

Lasse Liebst: Conceptualization (Lead), Formal Analysis (Lead), Supervision (Lead), Writing – original draft (Lead), Writing – review & editing (Lead), Methodology (Equal)

Lasse Baggesen: Investigation (Equal), Methodology (Equal), Writing – review & editing (Supporting)

Kasper Lykke Dausel: Investigation (Equal), Methodology (Equal), Writing – review & editing (Supporting)

Marie Rosenkrantz Lindegaard: Conceptualization (Supporting), Writing – review & editing (Supporting)

ORCID ID

Lasse Suonperä Liebst  <https://orcid.org/0000-0003-1062-2447>

Lasse Baggesen  <https://orcid.org/0009-0009-3721-695X>

Kasper Lykke Dausel  <https://orcid.org/0000-0001-5279-9000>

Marie Rosenkrantz Lindegaard  <https://orcid.org/0000-0002-1630-774X>

REFERENCES

- Agbo-Ajala, O., Viriri, S., Oloko-Oba, M., Ekundayo, O., & Heymann, R. (2022). Apparent age prediction from faces: A survey of modern approaches. *Frontiers in Big Data*, 5, 1025806. <https://doi.org/10.3389/fdata.2022.1025806>
- Alarifi, J., Fry, J., Dancey, D., & Yap, M. H. (2019). Understanding face age estimation: Humans and machine. In *Proceedings of the 2019 International Conference on Computer, Information and Telecommunication Systems (CITS)* (pp. 1–5). IEEE.
- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49(3–4), 227–266. <https://doi.org/10.1163/156853974X00534>
- American Sociological Association. (2018). American Sociological Association's code of ethics.
- Anda, F., Lillis, D., Le-Khac, N.-A., & Scanlon, M. (2018). Evaluating automated facial age estimation techniques for digital forensics. In *Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW)* (pp. 129–139). IEEE.
- Awad, D., Clifford, C. W. G., White, D., & Mareschal, I. (2020). Asymmetric contextual effects in age perception. *Royal Society Open Science*, 7(12), 200936. <https://doi.org/10.1098/rsos.200936>
- Brunswik, E. (1952). *The conceptual framework of psychology* (International Encyclopedia of Unified Science, Vol. 1, No. 10). University of Chicago Press.
- Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, 4(1), 5–6. <https://doi.org/10.1111/j.1745-6924.2009.01091.x>
- Clifford, C. W. G., Watson, T. L., & White, D. (2018). Two sources of bias explain errors in facial age estimation. *Royal Society Open Science*, 5(10), 180841. <https://doi.org/10.1098/rsos.180841>
- Dawson, R. (2011). How significant is a boxplot outlier? *Journal of Statistics Education*, 19(2), 2. <https://doi.org/10.1080/10691898.2011.11889610>
- Dehqan, A., Scherer, R. C., Dashti, G., Ansari-Moghaddam, A., & Fanaie, S. (2012). The effects of aging on acoustic parameters of voice. *Folia Phoniatrica et Logopaedica*, 64(6), 265–270. <https://doi.org/10.1159/000343998>
- Diekmann, Y., Smith, D., Gerbault, P., Dyble, M., Page, A. E., Chaudhary, N., Migliano, A. B., & Thomas, M. G. (2017). Accurate age estimation in small-scale societies. *Proceedings of the National Academy of Sciences*, 114(31), 8205–8210. <https://doi.org/10.1073/pnas.1619583114>
- Fiske, S. T. (2018). *Social cognition: Selected works of Susan Fiske*. Routledge.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. Wiley.
- Ganel, T., Sofer, C., & Goodale, M. A. (2022). Biases in human perception of facial age are present and more exaggerated in current AI technology. *Scientific Reports*, 12(1), 22519. <https://doi.org/10.1038/s41598-022-27009-w>
- Goffman, E. (1956). *The presentation of self in everyday life*. University of Edinburgh, Social Sciences Research Centre.
- Goffman, E. (1971). *Relations in public: Microstudies of the public order*. Basic Books.
- Gosselt, J. F., Van Hoof, J. J., & De Jong, M. D. T. (2012). Why should I comply? Sellers' accounts for (non-)compliance with legal age limits for alcohol sales. *Substance Abuse Treatment, Prevention, and Policy*, 7(1), 5. <https://doi.org/10.1186/1747-597X-7-5>
- Han, H., Otto, C., & Jain, A. K. (2013). Age estimation from face images: Human vs. machine performance. In *Proceedings of the 2013 International Conference on Biometrics (ICB)* (pp. 1–8). IEEE.
- Hastie, R., & Dawes, R. M. (2009). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Sage.

- Jarvie, C., & Renaud, K. (2024). Online age verification: Government legislation, supplier responsabilization, and public perceptions. *Children*, 11(9), 1068. <https://doi.org/10.3390/children11091068>
- Jones, L. K., Jennings, B. M., Higgins, M. K., & De Waal, F. B. M. (2018). Ethological observations of social behavior in the operating room. *Proceedings of the National Academy of Sciences*, 115(29), 7575–7580. <https://doi.org/10.1073/pnas.1716883115>
- Kask, K., & Tael-Oeren, M. (2017). How young/old does one look: Sales personnel's and laypersons' estimation of young people's age. *Juridica International*, 25, 82. <https://doi.org/10.12697/JI.2017.25.09>
- Klein, D. (2018). Implementing a general framework for assessing interrater agreement in Stata. *The Stata Journal*, 18(4), 871–901. <https://doi.org/10.1177/1536867X1801800408>
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Liebst, L. S., Baggesen, L., Dausel, K. L., Pallante, V., & Lindegaard, M. R. (2023). Human observers are accurate in judging personal relationships in real-life settings: A methodological tool for human observational research. *Field Methods*. <https://doi.org/10.1177/1525822X231158888>
- Liebst, L. S., Ejbye-Ernst, P., Bruin, M., Thomas, J., & Lindegaard, M. R. (2022). Face-touching behaviour as a possible correlate of mask-wearing: A video observational study of public place incidents during the COVID-19 pandemic. *Transboundary and Emerging Diseases*, 69(3), 1319–1325. <https://doi.org/10.1111/tbed.14094>
- Lowry, W. E. (2020). Its written all over your face: The molecular and physiological consequences of aging skin. *Mechanisms of Ageing and Development*, 190, 111315. <https://doi.org/10.1016/j.mad.2020.111315>
- Marcum, C. S. (2013). Age differences in daily social activities. *Research on Aging*, 35(5), 612–640. <https://doi.org/10.1177/0164027512453468>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Mehmetoglu, M., & Jakobsen, T. G. (2017). *Applied statistics using Stata: A guide for the social sciences*. SAGE Publications.
- Menon, V., Grover, S., Gupta, S., Indu, P., Chacko, D., & Vidhukumar, K. (2025). A primer on reliability testing of a rating scale. *Indian Journal of Psychiatry*, 67(7), 725–729. https://doi.org/10.4103/indianjpsychiatry_584_25
- Moyse, E. (2014). Age estimation from faces and voices: A review. *Psychologica Belgica*, 54(3), 255–265. <https://doi.org/10.5334/pb.aq>
- Punyani, P., Gupta, R., & Kumar, A. (2018). A comparison study of face, gait and speech features for age estimation. In A. Kalam, S. Das, & K. Sharma (Eds.), *Advances in electronics, communication and computing* (pp. 325–331). Springer.
- Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology*, 23(1), 1–12. <https://doi.org/10.1002/acp.1442>
- Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, 138(1), 146–174. <https://doi.org/10.1037/a0025750>
- Santani, D., Biel, J.-I., Labhart, F., Truong, J., Landolt, S., Kuntsche, E., & Gatica-Perez, D. (2016). The night is young: Urban crowdsourcing of nightlife patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 427–438). ACM.
- Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: An ecological approach to brain and behavior research. *Perspectives on Psychological Science*, 14(5), 841–859. <https://doi.org/10.1177/1745691619856350>
- Sörqvist, P., & Eriksson, M. (2007). Effects of training on age estimation. *Applied Cognitive Psychology*, 21(1), 131–135. <https://doi.org/10.1002/acp.1271>
- Starr, E., & Goldfarb, B. (2020). Binned scatterplots: A simple tool to make research easier and better. *Strategic Management Journal*, 41(12), 2261–2274. <https://doi.org/10.1002/smj.3199>
- Thorley, C. (2021). How old was he? Disguises, age, and race impact upon age estimation accuracy. *Applied Cognitive Psychology*, 35(2), 460–472. <https://doi.org/10.1002/acp.3744>
- Thorley, C., Acton, B., Armstrong, J., Ford, S., & Gundry, M. (2022). Are estimates of faces' ages less accurate when they wear sunglasses or face masks and do these disguises make it harder to later

- recognise the faces when undisguised? *Cognitive Research: Principles and Implications*, 7(1), 1–12. <https://doi.org/10.1186/s41235-022-00370-0>
- Turner, J. H. (2010). *Theoretical principles of sociology, volume 2: Microdynamics*. Springer.
- Twigg, J. (2015). Dress and age. In J. Twigg & W. Martin (Eds.), *Routledge handbook of cultural gerontology* (pp. 171–178). Routledge.
- Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2022). Credibility beyond replicability: Improving the four validities in psychological science. *Current Directions in Psychological Science*, 31(2), 162–168. <https://doi.org/10.1177/09637214211067779>
- Voegeli, R., Schoop, R., Prestat-Marquis, E., Rawlings, A. V., Shackelford, T. K., & Fink, B. (2021). Differences between perceived age and chronological age in women: A multi-ethnic and multi-centre study. *International Journal of Cosmetic Science*, 43(5), 547–560. <https://doi.org/10.1111/ics.12727>
- Wang, L., Koch, D. D., Hill, W., & Abulafia, A. (2017). Pursuing perfection in intraocular lens calculations: III. Criteria for analyzing outcomes. *Journal of Cataract and Refractive Surgery*, 43(8), 999–1002. <https://doi.org/10.1016/j.jcrs.2017.08.003>
- Wolf, B. (1999). Vicarious functioning as a central process-characteristic of human behavior. <https://brunswiksociety.org/wp-content/uploads/2022/06/Essay4.pdf>
- Yoon, K.-J. (2004). Not just words: Korean social models and the use of honorifics. *Intercultural Pragmatics*, 1(2). <https://doi.org/10.1515/iprg.2004.1.2.189>
- Yuan, B., Wu, A., & Zheng, W.-S. (2018). Does a body image tell age? In *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 2142–2147). IEEE.