

Do Social Judgements Made From Facial Photographs Predict Those Made In-Person?

Quarterly Journal of Experimental
Psychology
1–7

© Experimental Psychology Society 2026



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17470218261434907

qjep.sagepub.com



Anthony J. Lee¹, Amy A. Z. Zhao^{2,3}, Morgan J. Sidari^{2,4}
and Brendan P. Zietsch²

Abstract

People form rapid social judgements of others based on facial appearance. However, research to date relies on highly standardised photographs. We examined whether social judgements made from photographs predict judgements by different raters after an in-person interaction with the photographed persons. In a ‘speed-dating’ study, 689 participants (344 males, 345 females) rated each other on several traits after a 3-min interaction. Participants also provided facial photographs, which were rated on the same traits by a separate sample of 356 raters. Physical traits, such as attractiveness and athleticism, showed strong correspondence between face-only and in-person ratings. However, we also found correspondence between some non-physical traits, including creativity and intelligence for males, and confidence and extraversion for females. Correspondence between face-only and in-person judgements are required if personality traits can be accurately inferred from faces or if faces have lasting effects on interpersonal perceptions. Our findings have important implications regarding the ecological validity of research based solely on facial photographs.

Keywords

faces, person perception, attractiveness, social evaluation, impression formation

Received: 29 July 2025; revised: 16 February 2026; accepted: 2 March 2026

Introduction

People make quick and automatic judgements about others based on their facial appearance (Willis & Todorov, 2006), and research suggests that these initial impressions can have important social implications (Olivola et al., 2014). For instance, people are more likely to vote for election candidates whose faces are perceived as more competent (Todorov et al., 2005), individuals with trustworthy-looking faces are more likely to be promoted to managerial positions (Linke et al., 2016), and defendants who appear untrustworthy are more likely to receive harsher penalties in courtroom decisions (Wilson & Rule, 2015). Collectively, the research literature consistently suggests that facial appearance influences real-world outcomes.

Given this potential impact, it is unsurprising that a large body of research has investigated social judgements of

faces. However, this research has typically relied on highly controlled facial images; these are frontal, neutral-expression photographs (similar to passport photos) with extraneous factors (e.g. hair, clothing, background) tightly controlled. Highly standardised images are used with the aim of improving the reliability of research; however, it can

¹Faculty of Natural Sciences, Division of Psychology, University of Stirling, Scotland

²Centre for Psychology and Evolution, School of Psychology, University of Queensland, Brisbane, Australia

³Vrije Universiteit Amsterdam, The Netherlands

⁴Queensland Institute of Medical Research, Australia

Corresponding Author:

Anthony J. Lee, Faculty of Natural Sciences, Division of Psychology, University of Stirling, Stirling FK9 4LA, Scotland.

Email: anthony.lee@stir.ac.uk

be problematic because it reduces the ecological validity (Satchell et al., 2023). As such, it remains unclear whether judgements based solely on facial photographs extend to evaluations made in everyday social interactions.

There are several reasons to suspect that social judgements based solely on facial images may not translate to everyday interactions. First, faces never occur in isolation; contextual information – such as a person’s styling, posture, or expression – can be indicative of personality (Hester & Hehman, 2023) and is likely to play a major role in social evaluations. Second, other attributes about a person, such as their body (Bjornsdottir et al., 2024; Hu & O’Toole, 2023) or voice (McAleer et al., 2014), are likely to contribute to social impressions but are mostly overlooked in research focusing on faces. Third, while people can make snap judgements based on facial cues, these impressions may be quickly overridden when more diagnostic information – such as after verbal communication or witnessing behavioural cues – becomes available. Altogether, this suggests that the broader applicability of facial impression research remains uncertain.

Two studies have explored how first impressions from a photograph may update after an in-person interaction. Gunaydin et al. (2017) found that likeability judgements from a facial photograph predict a later likeability assessment after meeting in-person. Similarly, Satchell (2019) reported some consistency between initial judgements made from facial photographs of personality traits (e.g. extraversion, conscientiousness, neuroticism) and those later made following an in-person interaction. Both studies use a within-subjects design where the same rater makes judgements of targets in both contexts (i.e. after viewing a photograph, and again after an in-person meeting). While this design explores how impressions may update when additional information from an in-person interaction becomes available, it does not assess whether face-based impressions generalise across different contexts and independent raters, which is a fundamental assumption of the literature purporting that facial morphology impacts real-world decision-making. Within-subjects designs are also susceptible to individual-level biases or memory effects, potentially leading to greater or even artificial correspondence between in-person and face-only social impressions.

Here, we assess whether social trait ratings made from facial images predict social judgements made in-person by different raters. If judgements made solely from facial images are applicable to in-person interactions, then we should expect a positive correspondence between the two.

Methods

In-Person Ratings

In-person social judgements were collected using a speed-dating paradigm. Participants in the speed-dating study were part of a broader project investigating attraction and partner

preferences running between 2010 and 2019. Only data collected between 2014 and 2019 were included in the current analysis as the variables of interest were only collected in these phases. The final dataset included 687 participants (341 males, 346 females) who were recruited through the participant pool at the University of Queensland for course credit ($M=19.56$ years, $SD=2.97$ years).

Upon arrival, participants were verbally informed that they would have short interactions with other participants of the opposite sex that were in the same testing session. The test sessions were set up so that participants sat opposite from each other. In cases where there were an unequal number of male and female participants, surplus participants sat on their own. Participants interacted with each partner for 3 min, during which they were free to discuss whatever they wished. After 3 min, a bell would ring, indicating to participants to conclude their interaction and answer questions about their partner.

Participants responded to items about their partner on iPads. Relevant to the current analysis, participants were asked to rate their partner on 11 traits; these traits were ambitious, anxious, athletic, attractive, confident, creative, extraverted, funny, intelligent, kind and understanding, and organised. These traits were chosen as they are relevant to theories of partner choice and attraction. Ratings were made on a 7-point scale (1=*well below average*, 7=*well above average*). Once ratings were complete, a randomly chosen sex rotated to a new partner, and the process repeated until all dyads had interacted and rated each other.

In addition to the in-person social trait ratings, participants provided a facial photograph in the testing session. Participants were asked to stand in front of a blank wall with a neutral facial expression and face the camera, similar to a passport photo. Photographs were taken using a Canon IXUS 700. To keep images consistent, we used the flash and no zoom along with consistent indoor lighting.

For further details of the procedures of the speed-dating study, including additional measures collected, please see Lee et al. (2020), Sidari et al. (2021), Zhao et al. (2023, 2024), Harper et al. (2024), and Wainwright et al. (2024).

Facial Photograph Ratings

A separate group of participants were recruited to rate the facial photographs collected during the speed-dating study. This included 478 participants recruited from the University of Stirling participant pool for course credit (163 males, 304 females, 11 nonbinary or preferred not to say; $M=20.84$ years, $SD=6.90$ years). Participants were randomly assigned to rate the facial images on 1 of the 11 traits with the exception of attractiveness which was collected prior to the other traits (see open science and transparency statement below for more detail). We aimed to collect a minimum of 10 raters per trait and target sex.

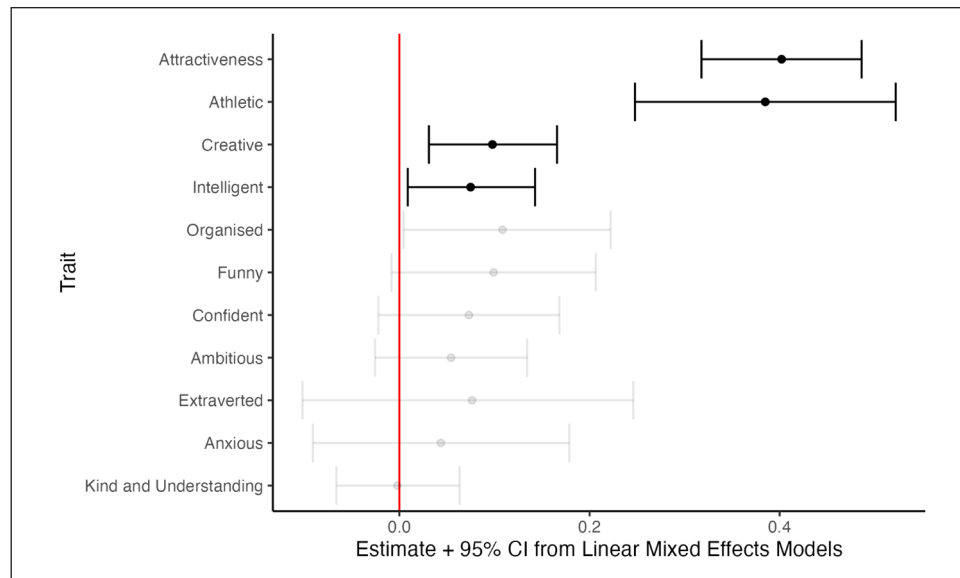


Figure 1. Estimated fixed effects from the linear mixed-effects models between mean trait ratings from facial images and in-person trait ratings for male targets.

Participants rated the facial images on their randomly assigned trait on a 9-point scale (1 = *very low in the trait*, 9 = *very high in the trait*). To match the in-person ratings from the speed-dating study, participants only rated the sex of the face that they were most attracted to. The faces were presented to participants in a randomised order.

Statistical Analysis

Data were analysed using linear mixed-effects models in the R software package (R Core Team, 2013) using the lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2015), and MuMIn (Bartoń, 2014) packages. Separate models were conducted for each trait and for each sex as we could expect the correspondence of trait judgements to differ by sex if people use different cues when considering male and female targets (Baudouin & Brochard, 2011). However, additional exploratory analyses that combine data from both sexes are included on the OSF (<https://osf.io/bjrgu/>). Mean trait ratings were calculated for each image on each trait; these were used as the predictors in the model with individual in-person ratings included as the outcome variable. Random effects were specified maximally following Barr et al. (2013), with grouping factors of rater, target, and test session included in the model.

Open Science and Transparency Statement

A pre-registration document is available on the OSF (<https://osf.io/resf6/>). The study was pre-registered prior to the online facial images being rated (with the exception of attractiveness ratings), but after data from the speed-dating study had been collected. The pre-registration document included the data collection procedures for the face-only

ratings, including how the sample size was determined, as well as the analysis plan and code for the reported statistical models. Attractiveness judgements were collected prior to pre-registration to help establish analytic procedures for the remaining traits (see pre-registration document for more detail). All data and analysis code supporting this article are available on the OSF (<https://osf.io/bjrgu/>).

Results

Below we report summaries of the estimated effects of mean trait ratings from facial images predicting in-person judgements from the linear mixed effects models. For full model results for each trait, including the Intraclass Correlations (ICCs), see the supporting documents on the OSF (<https://osf.io/bjrgu/>).

Male Targets

The estimated correspondence between in-person trait judgements and mean trait ratings based solely on the facial images for male targets are reported in Figure 1. There was a strong, significant correspondence between in-person and face-only ratings for the physical traits of ‘attractive’ and ‘athletic’ (marginal $R^2 = .11$ and $.09$ respectively). We also found a small, though significant correspondence for ‘creative’ and ‘intelligent’ (both marginal $R^2 = .01$). There was no significant correspondence between in-person and face-only judgements for any other trait.

Female Targets

Similar to the male targets, there was a strong, significant correspondence in the physical traits of ‘attractive’ and

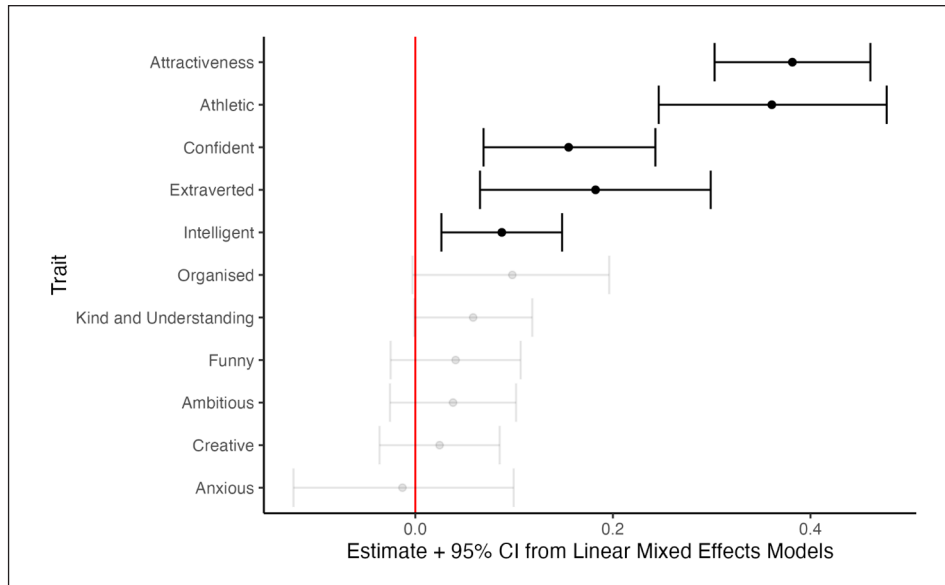


Figure 2. Estimated fixed effects from the linear mixed-effects models between mean trait ratings from facial images and in-person trait ratings for female targets.

‘athletic’ (marginal $R^2 = .11$ and $.09$, respectively), as well as a significant correspondence for ‘intelligent’ (marginal $R^2 = .01$). Additionally, we also found small though significant correspondences for traits ‘confident’ and ‘extraverted’ (marginal $R^2 = .02$ and $.03$, respectively). There was no other significant correspondence for the remaining traits (see Figure 2).

Discussion

For both male and female targets, the physical traits of attractiveness and athleticism showed strong face-image and in-person correspondence. This supports the idea that these visually salient traits can be reliably inferred from facial photographs alone. While this may seem intuitive, previous research has shown that variability in attractiveness ratings can be greater across different facial images of the same person than across photographs between individuals (Jenkins et al., 2011). This finding has been interpreted as evidence that attractiveness judgements of facial images reflect properties of the image rather than stable traits of the individual. However, our results directly challenge this interpretation: we find that attractiveness judgements from standardised facial images strongly relate to those made during face-to-face interactions, suggesting that such ratings do, in part, reflect stable characteristics of the target.

Interestingly, we also found a significant, positive correspondence for several non-physical traits. For male targets, we found significant correspondence between face-based and in-person judgements of intelligence and creativity. This finding is notable given the previous

research that has demonstrated some accuracy in judging intelligence based solely on facial information (Kleisner et al., 2014; Lee et al., 2017; Zebrowitz et al., 2002). Evolutionary psychology theories have proposed explanations for these effects; for instance, intelligence and creativity may be an indicator of underlying genetic quality (Haselton & Miller, 2006; Miller, 2000), which could also be reflected in physical attributes (Prokosch et al., 2005; Zebrowitz et al., 2002). Another proposed possibility is that intelligence and attractiveness may be genetically linked if intelligent individuals preferentially mate with facially attractive partners (Kanazawa & Kovar, 2004). However, attractiveness and intelligence has been found not to be correlated in a large, genetically informative sample (Mitchem et al., 2015), and we also do not find a significant correlation between intelligence and attractiveness for male targets in our sample, $r(342) = .01$, $p = .853$.

We also found significant correspondence for judgements of extraversion and confidence in female targets. This could be because more extraverted or confident women may signal these traits through increased use of cosmetics or facial adornments (Mileva et al., 2016). Alternatively, previous research has found that physically attractive women are perceived as more extraverted and confident (Brewer & Archer, 2007; Segal-Caspi et al., 2012). As such, any correspondence between face-only and in-person judgements for these traits could be driven by facial attractiveness. Indeed, further exploratory analysis found that for female facial images there was a strong correlation between attractiveness and extraversion ratings ($r[343] = .57$, $p < .001$) and between attractiveness and confidence ratings ($r[343] = .62$, $p < .001$).

However, despite finding significant correspondence for some non-physical traits, it should be noted that the estimated effects were small (marginal R^2 ranging from .01 to .03). This may indicate that while people can infer some traits from facial images alone, they only offer a limited glimpse into an individual's personality. Moreover, for many traits included in our study the correspondence was non-significant (albeit, the majority being in the positive direction), including some traits that have been the focus of substantial previous research (e.g. judgements of kindness in male faces; Stirrat & Perrett, 2010). For these non-significant traits, our findings therefore raise questions regarding the validity of claims that facial information meaningfully impacts real-world decision making. Such claims appear less plausible if the influence of facial appearance on social impressions is small or completely overridden once additional, more diagnostic information becomes available. While our results are limited to the specific traits assessed in our study, future research should consider examining a broader range of traits. Perhaps focusing on traits purported to be evolutionarily relevant such as dominance or trustworthiness (Oosterhof & Todorov, 2008) or primary traits such as the horizontal and vertical dimension from the recent meta-theory of social evaluation (Abele et al., 2021), may show a stronger face-to-in-person correspondence.

There are several caveats to consider when interpreting these results. First, as mentioned above, our study investigates between-rater correspondence effects; that is, participants making the in-person social judgement raters were different to the participants making the judgements based on the facial photographs. This design introduces additional individual-level variance, which could attenuate any potential face-to-in-person correspondence. However, previous research has suggested that there is good agreement between people for social trait ratings (Jones et al., 2021) and our approach mirrors the typical methodology of research in the field.

Second, the context in which in-person judgements were made differed from that of the face-only judgements. Research has suggested that context is important when making social evaluations (Satchell et al., 2023). In our study, the in-person ratings were made in a speed-dating context, which may encourage participants to evaluate others as a potential mate. As such, certain traits may be prioritised, weighted differently, or simply more relevant in this setting compared to that of the face-only evaluations. Such contextual differences contribute to the limited correspondence observed between in-person and face-only judgements for some traits.


Third, the sample of targets and raters included young university students from Western countries. Therefore, we cannot say to what extent these results would generalise to other populations, such as older or non-WEIRD individuals.

Finally, our research does not speak to the accuracy of social judgements, and indeed, prior research has questioned the accuracy of personality judgements based solely on facial images (Segal-Caspi et al., 2012). However, if people were able to accurately judge personality from faces, we would expect good agreement between face-based and in-person judgements. Similarly, if faces have a lasting influence on interpersonal impressions even in the presence of contextual information (e.g. expression, clothing, body language) and additional cues (e.g. body shape, voice), then we would expect consistent judgements across contexts. Given that we only observed such correspondence to a limited extent, our results raise questions about prior claims in the literature regarding the social impact of face-based impressions.

Future work could focus on the relative importance of multiple cues when forming social impressions. We could predict greater correspondence in impressions if we also consider the influence of additional cues (e.g. across different modalities) or contextual information (e.g. allowing for variation in facial expression, or the social context in which impressions are formed). Solely relying on facial photographs neglects these important sources of information, and our study raises questions about the ecological validity of earlier claims based solely on highly standardised, face-only stimuli. While there are niche situations where judgements are based solely or primarily on facial images (e.g. online dating), most real-world decisions, particularly those considered by previous studies, are not made in such isolation.

ORCID iDs

Anthony J. Lee  <https://orcid.org/0000-0001-8288-3393>

Amy A. Z. Zhao  <https://orcid.org/0000-0001-7366-8886>

Ethical Considerations

All procedures were approved by the University of Queensland Health and Behavioural Sciences, Low & Negligible Risk Ethics Sub-Committee and the University of Stirling's General University Ethics Panel.

Consent to Participate

All participants in the speed-dating study provided written consent to participate, while online participants provided electronic consent. Participants in the speed-dating study provided written consent for their images to be used in future studies.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement



All data and analysis code supporting this article are available on the OSF (<https://osf.io/bjrgu/>).

References

- Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Bartoń, K. (2014). *MuMIn: Multi-model inference*. <https://CRAN.R-project.org/package=MuMIn>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Baudouin, J., & Brochard, R. (2011). Gender-based prototype formation in face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 888–898.
- Bjornsdottir, R. T., Connor, P., & Rule, N. O. (2024). Social judgments from faces and bodies. *Journal of Personality and Social Psychology*, *127*(3), 455–468.
- Brewer, G., & Archer, J. (2007). What do people infer from facial attractiveness? *Journal of Evolutionary Psychology*, *4*(1), 39–49.
- Gunaydin, G., Selcuk, E., & Zayas, V. (2017). Impressions based on a portrait predict, 1-month later, impressions following a live interaction. *Social Psychological and Personality Science*, *8*(1), 36–44.
- Harper, K. T., Stanley, F., Sidari, M. J., Lee, A. J., & Zietsch, B. P. (2024). Role of accurate self-assessments in optimizing mate choice. *Personality and Social Psychology Bulletin*, *50*(4), 587–596.
- Haselton, M. G., & Miller, G. F. (2006). Women's fertility across the cycle increases the short-term attractiveness of creative intelligence. *Human Nature*, *17*(1), 50–73.
- Hester, N., & Hehman, E. (2023). Dress is a fundamental component of person perception. *Personality and Social Psychology Review*, *27*(4), 414–433.
- Hu, Y., & O'Toole, A. J. (2023). First impressions: Integrating faces and bodies in personality trait perception. *Cognition*, *231*, Article 105309.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–323.
- Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndakaihe, I. L. G., Bloxson, N. G., Lewis, S. C., Foroni, F., Willis, M. L., Cubillas, C. P., Vadiello, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., . . . Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*(1), 159–169. <https://doi.org/10.1038/s41562-020-01007-2>
- Kanazawa, S., & Kovar, J. L. (2004). Why beautiful people are more intelligent. *Intelligence*, *32*(3), 227–243.
- Kleisner, K., Chvatalova, V., & Flegr, J. (2014). Perceived intelligence is associated with measured intelligence in men but not women. *PLoS One*, *9*(3), Article e81237.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests for random and fixed effects for linear mixed effect models*. <https://CRAN.R-project.org/package=lmerTest>
- Lee, A. J., Hibbs, C., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2017). Assessing the accuracy of perceptions of intelligence based on heritable facial features. *Intelligence*, *64*, 1–8.
- Lee, A. J., Sidari, M. J., Murphy, S. C., Sherlock, J. M., & Zietsch, B. P. (2020). Sex differences in misperceptions of sexual interest can be explained by sociosexual orientation and men projecting their own interest onto women. *Psychological Science*, *31*(2), 184–192.
- Linke, L., Saribay, S. A., & Kleisner, K. (2016). Perceived trustworthiness is associated with position in a corporate hierarchy. *Personality and Individual Differences*, *99*, 22–27.
- McAlear, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS One*, *9*(3), Article e90779.
- Mileva, V. R., Jones, A. L., Russell, R., & Little, A. C. (2016). Sex differences in the perceived dominance and prestige of women with and without cosmetics. *Perception*, *45*(10), 1166–1183.
- Miller, G. (2000). *The mating mind*. Doubleday.
- Mitchell, D. G., Zietsch, B. P., Wright, M. J., Martin, N. G., Hewitt, J. K., & Keller, M. C. (2015). No relationship between intelligence and facial attractiveness in a large, genetically informative sample. *Evolution and Human Behavior*, *36*, 240–247.
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Science*, *18*(11), 566–570.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092.
- Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher g-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, *33*(2), 203–213.
- R Core Team. (2013). *A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Satchell, L. P. (2019). From photograph to face-to-face: Brief interactions change person and personality judgments. *Journal of Experimental Social Psychology*, *82*, 266–276.
- Satchell, L. P., Jaeger, B., Jones, A., López, B., & Schild, C. (2023). Beyond reliability in first impression research: Considering validity and the need to 'mix it up with folks'. *Credibility in Social and Personality Psychology*, *18*, Article e10211.
- Segal-Caspi, L., Roccas, S., & Sagiv, L. (2012). Don't judge a book by its cover, revisited: Perceived and reported traits and values of attractive women. *Psychological Science*, *23*(10), 1112–1116.
- Sidari, M. J., Lee, A. J., Murphy, S. C., Sherlock, J. M., Dixon, B. J. W., & Zietsch, B. P. (2021). Preference for sexually dimorphic body characteristics revealed in a large sample of

- speed daters. *Social Psychological and Personality Science*, *12*(2), 225–236.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.
- Wainwright, H. M., Zhao, A. A. Z., Sidari, M. J., Lee, A. J., Roberts, N., Makras, T., & Zietsch, B. P. (2024). Laughter and ratings of funniness in speed-dating do not support the fitness indicator hypothesis of humour. *Evolution and Human Behavior*, *45*(1), 75–81.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.
- Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331.
- Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, *28*(2), 238–249.
- Zhao, A. A. Z., Harrison, K., Holland, A., Wainwright, H. M., Ceccato, J., Sidari, M. J., Lee, A. J., & Zietsch, B. P. (2023). Objectively measured facial traits predict in-person evaluations of facial attractiveness and prosociality in speed-dating partners. *Evolution and Human Behavior*, *44*(4), 315–323.
- Zhao, A. A. Z., Sidari, M. J., Lee, A. J., & Zietsch, B. P. (2024). Speed-dating and simulation data explain the discrepancy between stated and revealed mate preferences. *European Journal of Personality*, *39*(4), 535–557.