Journal of Economic Behavior and Organization 218 (2024) 296–308

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research Paper

How beautiful people see the world: Cooperativeness judgments of and by beautiful people ☆

Adam Zylbersztejn [a,b], Zakaria Babutsidze [c], Nobuyuki Hanaki [d,e], Astrid Hopfensitz [f,*]

[a] Univ Lyon 2, Université Lumière Lyon 2, GATE UMR 5824, 69130 Ecully, France
[b] Research Fellow at Vistula University Warsaw (AFiBV), Warsaw, Poland
[c] SKEMA Business School, Université Côte d'Azur (GREDEG) and OFCE, Sciences Po Paris, France
[d] Institute of Social and Economic Research, Osaka University, Japan
[e] University of Limassol, Cyprus
[f] Emlyon Business School, GATE UMR 5824, F-69130 Ecully, France

A R T I C L E I N F O

JEL classification:
C72
D83

Keywords:
Cooperation
Beauty
Perception
Hidden action game
Experiment

A B S T R A C T

Perceived beauty is one of the strongest predictors of perceived cooperativeness, causing the "beauty bias". Through a large three-stage incentivized behavioral and rating experiment (N = 357), we study (1) the beauty bias in incentivized predictions of cooperativeness and (2) the ex-post relevance of beauty ratings for predicting cooperativeness in an incentivized game. We additionally (3) investigate if one's beauty influences the beauty bias in predictions of cooperativeness of others. Our findings demonstrate the robustness of the beauty bias and its irrelevance for making accurate predictions. We further observe that individuals are affected by the beauty bias irrespective of their beauty. Overall, the results highlight the importance of strong institutions that protect individuals from falling prey to the beauty bias.

1. Introduction

Attractiveness has a persistent and long-term importance not only for labor market earnings (Scholz and Sicinski, 2015), but for economic success in general (Hamermesh, 2013). Such a beauty premium exists for both males and females (e.g., Hamermesh and Biddle, 1994; Doorley and Sierminska, 2015) and across many different cultures (Bokek-Cohen and Davidovich, 2011; Cunningham et al., 1995). Beautiful people benefit from an attractiveness halo effect (Dion et al., 1972) that leads them to be perceived as more intelligent (Zebrowitz et al., 2002) and healthier (Kalick et al., 1998), as being better leaders (Budesheim and DePaola, 1994), and as being more trustworthy (Wilson and Eckel, 2006). However, the latter study also argues that beautiful people are more likely to trigger disappointment since they do not live up to the high expectations others put into them.

☆ We acknowledge financial support from Université Lyon 2 through APPI call (project PEROSA). This research has been performed within the framework of LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) run by the French National Research Agency (ANR). Imen Bouhlel, Mathieu Chevrier and Marie-Sophie Roul provided quality research assistance. Quentin Thévenet provided valuable assistance with software programming. We also thank Agnès Festré for her support of the project. Adam Zylbersztejn acknowledges VISTULA Fellowship. Nobuyuki Hanaki acknowledges various supports from the Joint Usage/Research Center at the Institute of Social and Economic Research, Osaka University.

* Corresponding author.
E-mail address: hopfensitz@em-lyon.com (A. Hopfensitz).

https://doi.org/10.1016/j.jebo.2023.12.020
Received 19 June 2023; Received in revised form 6 December 2023; Accepted 11 December 2023

Available online 20 December 2023
0167-2681/

While there is much research on how beautiful people are perceived, there are fewer studies on the distinct features of the decision-making of the beautiful (Teng et al., 2022). While many lay people might believe that beauty is in the "eye of the beholder", there is a strong degree of consistency in the beholders' perceptions of beauty (Hamermesh and Biddle, 1994). Hence, in this paper we consider as beautiful those who are perceived as beautiful by others (see, e.g., Belot et al., 2012, for a similar approach). The present paper aims at adding to our understanding of how one's beauty relates to their behavior and perception of the world. We do this in three steps. We first elicit the willingness to engage in cooperative behavior in general from a set of individuals. We then replicate the finding that individuals perceived as more beautiful are also perceived as more likely to cooperate (Hypothesis 1). We then investigate if one's beauty influences general beliefs regarding cooperativeness of others (Hypothesis 2) and if one's beauty affects their reaction to the beauty of others in social interactions (Hypothesis 3). Jointly, this enables us to understand if beautiful people are indeed different regarding (i) willingness to cooperate with others and (ii) ways of judging others' cooperativeness.

An important body of research focuses on the accuracy of the "what is beautiful is good" heuristics (Dion et al., 1972). Beauty might be a useful proxy to follow under certain circumstances for example in labor markets due to its complex implications (Hamermesh, 2013; Altman et al., 2021). Its predictive power of personal attributes has meanwhile been mostly refuted: e.g., for intelligence (Deryugina and Shurchkov, 2015; Mitchem et al., 2015; Scholz and Sicinski, 2015), and trustworthiness (Solnick and Schweitzer, 1999; Wilson and Eckel, 2006; Van Leeuwen et al., 2018) – though some theoretical arguments for their persistence exist (Symons, 1979; Thornhill and Gangestad, 1999). Whether beauty is predictive of health is still a highly contentious issue (Kalick et al., 1998; Rhodes et al., 2001). For simplicity and consistency with the existing literature, we will refer to a beauty "bias", although under certain circumstances it could also be a beauty "heuristic". For example, certain elements of beauty (like smooth skin or white teeth) are clearly correlated with health states. In addition, many of the individual characteristics of beauty are also correlated with age which constitutes another factor influencing both health and beauty. The theoretical arguments have mainly focused on the reproductive advantage of certain physical traits considered beautiful in females, notably the waist-to-hip ratio (see, e.g. Bovet, 2019). Empirical support for this relationship has been indeed observed for females but not for males (Weeden and Sabini, 2005). Consequently, beauty in males and females may lead to very different reactions that additionally depend on the raters' gender – for instance, beauty in females may trigger price discrimination by males (Ruffle et al., 2022). When analyzing ratings and perceptions of beauty, it is therefore important to focus as well on the gender of the rated but also on the gender of the rater.

Given the privileged way in which they are perceived by others, beautiful people are also likely to have distinct experiences. Social stereotypes can lead to self-fulfilling prophecies in a number of different situations (Snyder et al., 1977). For instance, the beautiful may not be better leaders per se, but may be preferred for taking such positions and thus gain valuable expertise in leadership. In the same vein, one may conjecture that the fact that beautiful people attract more trust implies that they also get more opportunities to learn about the benefits of cooperating with others, but also that they may be less inclined to engage into singular cooperation due to the abundance of opportunities.

Recent empirical research aims at distinguishing between the "what is beautiful is good" idea from social psychology that would predict beautiful individuals to be selfless and trustworthy and the "evolutionary entitlement" hypothesis that would predict selfish behavior by beautiful individuals due to a sense of entitlement (Teng et al., 2022). Teng et al. (2022) find no evidence that beauty is correlated with more selfless behavior, but with an important caveat: the key measure of beauty is self-perceived attractiveness. Earlier studies based on external beauty ratings provide mixed evidence: more beautiful individuals may be less cooperative (Wilson and Eckel, 2006) or not (Belot et al., 2012). We add to this literature by linking external ratings of beauty, and behavior in an incentivized hidden action game with an anonymous partner which constitutes a classic setting for studying cooperation in a principal-agent relationship with moral hazard.[1] Our results indicate that beauty falls short of predicting actual behavior. However even though beauty is non-informative regarding cooperativeness, we observe a strong beauty bias with respect to ratings and predictions of cooperativeness.

More generally, our study is also related to the literature on how implicit biases can be overcome (Devine, 1989). It has been shown that biases (e.g., with respect to gender or weight, and usually measured through the implicit association test) are hard (though not impossible) to overcome. For instance, in the case of racial biases, the exposure to counter-stereotypical exemplars (e.g., Dasgupta and Rivera, 2008) or the activation of egalitarian goals (Legault et al., 2011; Mann and Kawakami, 2012) may reduce implicit biases. Our study focuses on whether benefiting from a certain bias throughout one's lifetime affects one's susceptibility to the same bias when interacting with others. In the domain of gender biases, for example, it has been observed that exposure to sexism might reduce the bias (Ramos et al., 2016). Though self-perceived attractiveness often deviates from external ratings, the two measures remain correlated (Feingold, 1992). Beautiful individuals might thus be aware of being preferentially treated by others, or might have developed higher self-esteem thanks to their beauty (Agthe et al., 2011). Herein, we investigate to which degree one's own beauty moderates the beauty bias one displays toward other people. We find no support for the existence of this source of heterogeneity in the beauty bias which points to the generality of this judgment error.

2. Methods

We conducted an experiment consisting of three stages with a total of $N = 357$ subjects in three experimental economic laboratories in France: LEEP in Paris (Stage 1 carried out in October 2019 with $N = 76$ participants), GATE-LAB in Lyon (Stage 2 carried

[1] Note that the participants are not aware of their interaction partner's beauty while making decisions in the game. Thus, the game provides a general measure of cooperativeness.

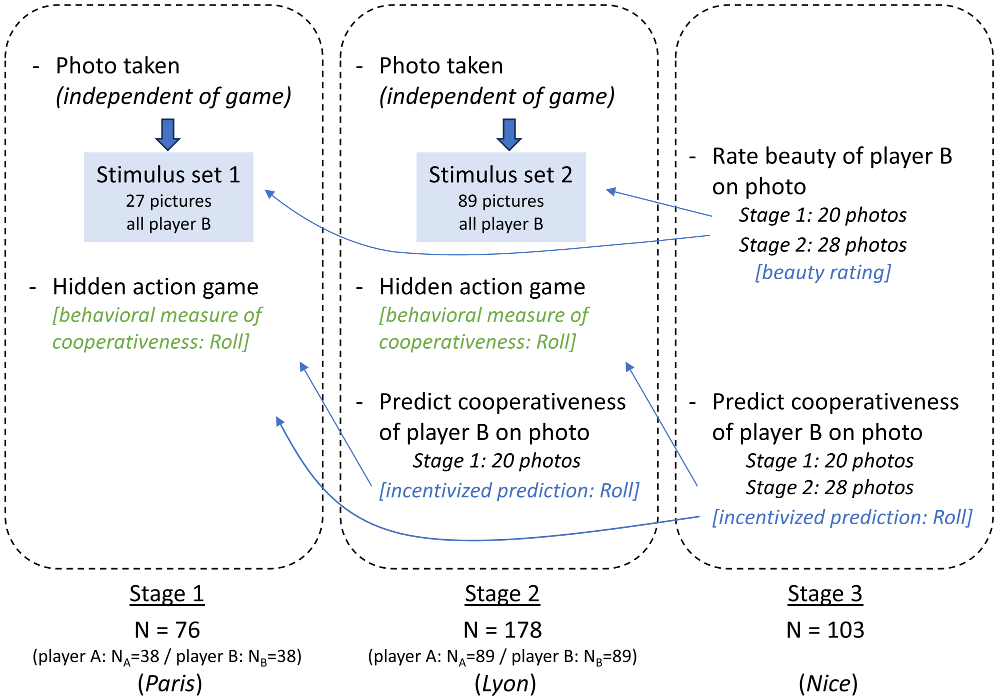

Fig. 1. Overview of the three experimental stages. Note that the hidden action game was in Stages 1 and 2 played with another (anonymous) individual from the same stage. Predictions regarding cooperativeness in the hidden action game were always made by individuals from a later stage. Numbers of observations (N) in the figure refer to the participants in each stage. Due to the stratification of stimuli in Stages 2 and 3, solely 27 out of 38 player Bs from Stage 1 were kept for later stages (see Section 2 for details).

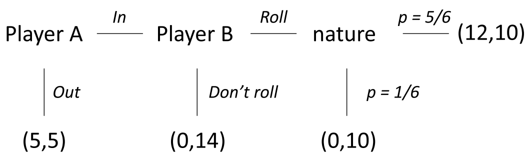

Fig. 2. Hidden action game. Cooperativeness is measured by the target player choosing $Roll$.

out between May and September 2022, $N = 178$) and LEEN in Nice (Stage 3 carried out in April 2023, $N = 103$). Fig. 1 provides a visual summary of the three experimental stages. Stage 1 was an initial stage to collect both behavioral observations and photographs from a first set of participants. Stage 2 is our main stage of interest, where participants both evaluate the cooperativeness of Stage 1 participants and provide their own photographs and behavior. Participants in Stage 3, predict again cooperativeness of all previous participants (Stage 1 and Stage 2) and also provide beauty ratings for participants from Stages 1 and 2. We will provide details on the procedures of each stage below.

The experiment was fully computerized: subjects were recruited using ORSEE (Greiner, 2015), and all the experimental tasks were programmed with z-Tree (Fischbacher, 2007).[2]

Stage 1 (Paris): initial set of stimuli collection. Participants were assigned to the role of player A or player B in a one-shot hidden action game (Charness and Dufwenberg, 2006) presented in Fig. 2. Player A could either go *Out* yielding 5 EUR to each player and ending the interaction, or go *In* letting player B decide about the outcome of their interaction. Player B could either choose to *Roll* a die (yielding 10 EUR to herself with certainty and exposing player A to a lottery: 12 EUR with 5/6 probability vs. 0 EUR with 1/6 probability), or not to *Roll* (yielding 14 EUR to herself and 0 EUR to player A with certainty).[3]

Like Charness and Dufwenberg (2006), we used the strategy method and simultaneously elicited both players' decisions: player B made a decision without knowing the decision by player A, and that decision was only implemented if player A had chosen *In*. Hereafter, we focus on the behavior of player B and use their decision in the hidden action game as the main incentivized measure of cooperativeness.

[2] ORSEE is a web-based Online Recruitment System, specifically designed for organizing economic experiments. In each location, the subject pool consists of registered volunteers, predominantly students from local campuses.

[3] In addition to experimental payoffs, each participant in each of the three experimental stages received a show-up fee of 5 EUR.

We also collected photographs of all participants. Upon arrival to the laboratory and prior to learning about the rules of the hidden action game (specific instructions are provided as Part 1 of Appendix A), each participant was invited to a separate room for a portrait picture to be taken. Portraits were taken in front of a neutral background and were later cropped and re-centered so as to contain only the head and shoulders of the participant. Photographs were later linked to behavior in the hidden action game and served as visual stimuli in the two subsequent experimental stages. At the end of the stage (this also applies to stages 2 and 3), we elicited basic socio-demographic variables related to age, gender, and education.

Out of the set of 38 pictures presenting player Bs, we ultimately selected a subset of 27 pictures to be used in the next experimental stage (63% female, average age 22.8 with SD 2.8, 48% rolled a die). For details see also the stratification exercise described below.

Stage 2 (Lyon): prediction task with stimuli from Stage 1 (Paris) and collection of a second set of stimuli. In the second experimental stage, participants performed multiple tasks, each of which generated payoffs and the rules of which were presented on a rolling basis. Each task was incentivized. First, participants played the hidden action game in exactly the same way as in Stage 1. Second, they made a series of twenty predictions regarding behavior of player Bs from Stage 1 (i.e., whether that player decided to $Roll$ or not). To calibrate prior beliefs about the cooperativeness of these player Bs, we informed the participants that the rate of choosing $Roll$ in the series was 50%, but we did not disclose the length of the series of pictures (see Vogt et al., 2013, for a similar design). Prior to making a prediction, participants saw a picture of the target player B from Paris. Each participants saw 20 pictures out of the total stimulus set of 27. Each picture was displayed on the screen for 5 seconds after which the participant could press a button and move on to the prediction-making stage. Sets of stimuli were further standardized across participants with respect to target players' individual characteristics: ethnicity and gender. The visual stimuli were stratified for each participant according to the target player B's gender and visual ethnicity (as evaluated by the authors themselves based on the pictures). In each series of stimuli, we made sure that for each gender and ethnicity present in the set, equally many players Bs chose and did not choose to $Roll$. To the extent that perceived cooperativeness may be correlated with these individual characteristics, through this design we make sure that the variation in predictions in Stage 2 is not a mere artifact due to stimuli sampling differences. The prediction task was incentivized: a correct prediction resulted in a payment of 10 euros while a wrong prediction gave a payment of 2 euros. At the end of the stage, one of the twenty predictions was randomly selected for payout. No feedback was provided while performing the task.[4]

As in Stage 1, we also collected photographs from all participants. Prior to completing the experimental tasks and receiving any further instructions, each participant was invited to have their portrait taken. Pictures were then processed like those from Stage 1 to provide a comparable set of visual stimuli. Once again, we focus on participants acting as player B in the hidden action game. The resulting sample consists of 89 individuals (54% female, average age 22.6 with SD 4.3, 53% choose $Roll$).

Stage 3 (Nice): Beauty ratings and cooperativeness predictions of target players from Stage 1 (Paris) and Stage 2 (Lyon). Our last stage consisted of several parts, the rules of which were revealed to participants as the stage unfolded. In the first part participants (64.1% females, average age 23.5 with SD 7.5) were presented with the rules of the hidden action game implemented in the two previous experimental stages and then asked to rate a series of portraits drawn from those stages (see Appendix B). Portraits were presented on the left-hand side of the screen, and the ratings were entered on the right. For each stimulus, two consecutive sets of four ratings were to be made. There was no time limit for making these ratings.

The first screen contained our main incentivized measure of perceived trustworthiness: the predicting whether the player B in the picture decided to $Roll$ in the hidden action game or not. The incentive structure was the same as in Stage 2 (i.e., one randomly selected prediction was payoff-relevant, giving 10 euros for a correct prediction and 2 euros in case of an incorrect prediction). We also fixed priors at 50% as in Stage 2.

Besides the prediction about behavior, we also asked participants to rate individuals presented in the stimuli along seven dimensions. The ratings were done on two consecutive screens and were not incentivized. Specifically, ratings on screen 1 concerned: beauty, trustworthiness, and intelligence; and ratings on screen 2 concerned: self-confidence, masculinity, competence, and predictability. Participants gave their answers on a seven-point Likert scale.[5]

The ratings of portraits were split across two parts. During the first part portraits from Stage 2 (Lyon) were rated, and in the second part those from Stage 1 (Paris) were rated. In the first part, we presented participants with 28 photo stimuli drawn from the photographs collected during Stage 2. To standardize the sequence across raters and guarantee a meaningful number of individual ratings per stimuli, each set of 28 stimuli was balanced in a 2×2 manner: males and females were equally represented, and for each gender the $Roll$ rate was 50%. As a result, we obtained between 19 and 51 ratings per stimulus (on average 31.8, SD 6.3).

[4] Before and after the prediction task we also elicited other variables that are related to a different pre-registered study (registration #35682 at AsPredicted.org) and are not part of the analyses reported in this paper. After receiving specific paper instructions (see Appendix A) and answering a short comprehension quiz, but prior to viewing the first stimuli in a series, participants provided a certainty equivalent they would be willing to accept instead of a payoff depending on a randomly chosen prediction. This is done through a standard Becker-DeGroot-Marschak mechanism (Becker et al., 1964) which is an incentive-compatible procedure for measuring willingness to accept (WTA) used herein, as well as its alternative: willingness to pay (WTP). Implementation details are provided in Appendix A. We also measured several individual characteristics and abilities: emotional intelligence captured by "Reading the Mind in the Eyes" test score (Baron-Cohen et al., 2001), capacity to apply backward induction in the Race-to-17 game (Gneezy et al., 2010; Chen et al., 2019), cognitive skills as measured by the standard 3-item Cognitive Reflection Test (Frederick, 2005), two questions on general attitudes towards trusting other people and trusting strangers adopted from the German Socio-Economic Panel Study, and an incentivized lottery task for measuring risk preferences (Gneezy and Potters, 1997).

[5] The specific terms and orientation of the scales was: "beautiful 1 … 7 not beautiful", "trustworthy 1 … 7 untrustworthy", "little intelligent 1 … 7 very intelligent", "shy 1 … 7 self-confident", "very feminine 1 … 7 very masculine", "little competent 1 … 7 very competent" and "predictable 1 … 7 unpredictable". For the analysis, we aligned scales (i.e., inverted the scales regarding beauty, trustworthiness and predictability) such that for all ratings higher values imply a stronger perception of the characteristic relative to its opposite.

Table 1
Aggregate mean ratings: gender differences.

Rating:	Overall	Rater			Stimulus		
		Male	Female	p	Male	Female	p
Prediction *Roll* (0/1)	0.532	0.520	0.537	0.538	0.524	0.538	0.446
Beautiful	4.432	4.098	4.608	<0.001	4.001	4.770	<0.001
Trustworthy	4.305	4.210	4.356	0.004	4.058	4.500	<0.001
Intelligent	4.705	4.517	4.804	<0.001	4.674	4.729	0.522
Self-confident	3.960	3.926	3.981	0.298	4.082	3.866	0.069
Masculine	4.000	3.949	4.031	0.020	5.665	2.695	<0.001
Competent	4.645	4.427	4.761	<0.001	4.661	4.632	0.636
Predictable	3.934	4.075	3.847	<0.001	3.775	4.059	<0.001

Note. Pooled stimuli ($N = 116$) collected in Lyon and in Paris. Mean ratings by male vs. female raters from Nice are matched at the stimulus level, p-values correspond to comparisons of distributions based on two-sided signrank test. Mean ratings for male ($N = 51$) vs. female ($N = 65$) player Bs are compared using ranksum test. Each of the seven characteristics (Beautiful, Trustworthy, Intelligent, Self-confident, Masculine, Competent, Predictable) is measured on a 1 to 7 Likert scale (with higher values implying stronger perception of a characteristic relative to its opposite).

In the second part of the stage, we informed the participants that they were about to repeat the same task with a different set of pictures. As before, we did not disclose the exact length of the sequence, but we emphasized that the number of items was not the same as in the previous task. The implementation procedure – the source of stimuli and their stratification, as well as the incentive structure of the prediction task – followed the one in Stage 2. Stimuli in this part came from Stage 1. This exercise yielded between 24 and 77 ratings per stimulus (on average 57.1, SD 18).[6]

3. Ethics and pre-registered hypotheses

All participants in Stages 1 and 2 were informed in the invitation email that during the experiment they would be photographed and that these photos would be used for strictly scientific purposes within this research project. Participants were asked to only accept the invitation and consequently to participate in the experiment if they agreed with these conditions. Additionally, upon arrival at the laboratory and before any picture was taken, all participants gave their written consent for the production and the strictly scientific usage of their pictures. This protocol was approved by the GATE-Lab Review Board (#2020-10).

All hypotheses and their empirical testing procedures were pre-registered at AsPredicted.org (registration #119687).[7] Specifically, we wanted to evaluate whether individuals perceived as beautiful are also expected to be more cooperative (Hypothesis 1). Supporting evidence is provided in Section 4.2. We further tested whether these individuals have higher expectations about the cooperativeness of others (Hypothesis 2), and whether there is a difference between the less and the more beautiful individuals in reaction to the beauty in others when judging their cooperativeness (Hypothesis 3). The evidence regarding these hypotheses can be found in Section 4.3. We also pre-registered our investigation of the predictive role of individual characteristics (i.e., age and gender) and their use as control variables (see Section 4.1). The analyses are based on standard correlation measurements and parametric regression models.

4. Results

Our empirical investigation focuses on the relationship between beauty (both one's own and that of the person seen in the stimulus) and the (perceived) cooperativeness of the person in the picture. We capture perceived cooperativeness through the prediction of whether the target player chose *Roll* in the hidden action game. As the hidden action game is a game of moral hazard, participants (i.e., agents) that choose *Roll* in this game are clearly performing a cooperative act with respect to their principal (Charness and Dufwenberg, 2006).

In what follows, we first present raters' assessment of the target subject's characteristics, with a particular focus on gender differences (both across raters and across stimuli). Then we investigate which characteristics are truly predictive of cooperative behavior, and which characteristics individuals take into account in their cooperativeness predictions. Finally, we separately investigate the dependence of the beauty bias on the beauty of the rater.

4.1. Perception of stimuli: exploring gender-based heterogeneity

The first column of Table 1 summarizes the aggregate mean ratings of the stimuli set of pictures presenting targets (i.e., player Bs) from Lyon ($N = 89$) and from Paris ($N = 27$). The correlation structure across the ratings of individual characteristics is summarized in Table 7 of Appendix C. In terms of our main incentivized measure of perceived cooperativeness, i.e., predicting whether a given player B rolled a die or not, raters happen to be overly optimistic: the likelihood of making a prediction of *Roll* is 0.532 which significantly differs from the benchmark value of 0.5 ($p < 0.001$, two-sided t-test).

[6] These figures are based on data from 77 raters. Data from 25 raters were lost due to a computer glitch during one of the sessions.

[7] The pre-registration is available at: https://aspredicted.org/P7J_Z13.

Table 2
Stimulus-level determinants of male rater–female rater gap: regression analysis.

	coeff (SE)	p	coeff (SE)	p	coeff (SE)	p	coeff (SE)	p
Rating:	Prediction *Roll*		Beautiful		Trustworthy		Intelligent	
Intercept	-0.035 (0.040)	0.508	-0.145 (0.117)	0.215	0.155 (0.107)	0.150	-0.264 (0.092)	0.005
$1[Lyon]$	0.007 (0.039)	0.381	-0.143 (0.114)	0.212	-0.125 (0.105)	0.233	0.027 (0.090)	0.768
$1[Female\_stimulus]$	0.022 (0.033)	0.851	-0.456 (0.097)	<0.001	-0.366 (0.089)	<0.001	-0.079 (0.076)	0.303
R^2	0.004		0.169		0.135		0.011	
$Prob > F$	0.795		<0.001		<0.001		0.548	
Rating:	Self-confident		Masculine		Competent		Predictable	
Intercept	0.179 (0.118)	0.133	-0.303 (0.092)	0.001	-0.302 (0.099)	0.003	0.191 (0.131)	0.147
$1[Lyon]$	-0.256 (0.116)	0.029	0.028 (0.090)	0.753	0.068 (0.096)	0.483	0.103 (0.128)	0.421
$1[Female\_stimulus]$	-0.066 (0.098)	0.504	0.356 (0.077)	<0.001	-0.151 (0.082)	0.069	-0.076 (0.109)	0.485
R^2	0.044		0.160		0.035		0.011	
$Prob > F$	0.080		<0.001		0.135		0.541	

Note. Dependent variable is the stimuli-level difference between male rater vs. female rater mean scores. Pooled stimuli ($N = 116$) collected in Lyon ($1[Lyon] = 1$) and in Paris ($1[Lyon] = 0$). $1[Female\_stimuli] = 1$ if the player B seen in the stimuli is female, and $= 0$ otherwise. Each of the seven characteristics (Beautiful, Trustworthy, Intelligent, Self-confident, Masculine, Competent, Predictable) is measured on a 1 to 7 Likert scale (with higher values implying stronger agreement with the characteristic).

The middle section of the table presents ratings separately for male and female raters. The values in the first row indicate that the bias of the female raters is not significantly different from the bias of male raters (p =0.538, signrank test). Regarding the seven evaluations of individual characteristics, we observe rater gender differences in six out of seven cases. Male raters perceive player Bs to be less beautiful, less trustworthy, less intelligent, less masculine, less competent, and more predictable.

The right section of Table 1 breaks down the aggregate mean ratings according to the gender of the player B seen in the picture. Overall, we find that males are perceived as less beautiful, less trustworthy, less feminine, and less predictable.

From the above, we can thus conclude that gender differences in rating arise on both the rater side and the side of the rated. It might thus be possible that the two interact. In Table 2 we regress the observed stimuli-level rater-gender gap on stimuli-specific variables: the gender of the player B seen in the picture and the location of the experiment (i.e., Lyon vs. Paris). This analysis confirms the previously described stability of incentivized predictions of player B's decision to *Roll*: ratings are the same for male and female raters and do not depend on the player B's gender nor on their location. We further observe that player B's gender contributes to the rater-gender gap for beauty, trustworthiness and masculinity. In particular, in terms of beauty male raters assign lower scores than female raters if a stimuli presents a female. No significant difference arises when a stimuli presents a male. A similar effect is observed for perceived trustworthiness. Regarding masculinity, male raters give lower ratings than female raters to males, while no such difference arises for female stimuli. Finally, in terms of intelligence and competence we observe that male raters systematically assign lower ratings than female raters regardless of player B's gender.

4.2. What makes a person (appear to be) cooperative?

We now turn to investigating which characteristics are related to cooperativeness, and which characteristics are used by raters to predict a person's cooperativeness.

The first question we ask in this part is: to what extent does the aforementioned set of characteristics predict cooperativeness? To answer this question, in Table 3 we present results of a logistic regression of the actual decision made by the player B (*Roll* or not) on the individual ratings outlined in the previous section. We further control for the location of the experiment and player B's (self-reported) age.[8] The result is clear: no explanatory variable reaches significance at the conventional 5% level.[9]

Thus, the next question that arises is the following: do raters nonetheless condition their perception of cooperativeness on these individual characteristics?

[8] For a target player B in stimulus i, the probability of decision $Roll$ is modeled as follows: $Pr(Roll_i) = F(\alpha_0 + \alpha_1 \times Beautiful_i + \alpha_2 \times Trustworthy_i + \alpha_3 \times Intelligent_i + \alpha_4 + Self-confident_i + \alpha_5 \times Masculine_i + \alpha_6 \times Competent_i + \alpha_7 \times Predictable_i + \alpha_8 \times 1[Female\_stimulus_i] + \alpha_9 \times 1[Lyon_i] + \alpha_{10} \times Age_i$ where F is a standard logistic cumulative distribution function.

[9] Appendix C provides complementary evidence on the corresponding effect sizes. Comparing individuals that rolled to those that did not, we find that for all seven characteristics of interest Cohen's $d < 0.2$, i.e., below the level conventionally considered as small (see Table 6). Finally, even though the coefficient for females reaches a significance level below 10 percent in the regression, our results do not point to a gender difference regarding cooperative behavior. In aggregate terms, there is no evidence for gender difference in cooperativeness: the rates of $Roll$ of males vs. females are 0.529 vs. 0.508 (proportion test: $p = 0.816$).

Table 3
Predictors of cooperativeness: logistic regression analysis.

	coeff (SE)	p
Intercept	-11.767 (6.467)	0.069
Beautiful	-0.052 (0.588)	0.930
Trustworthy	0.126 (0.727)	0.863
Intelligent	0.862 (1.106)	0.436
Self-confident	0.086 (0.454)	0.850
Masculine	0.936 (0.528)	0.076
Competent	-0.096 (1.300)	0.941
Predictable	-0.177 (0.776)	0.820
$1[Female\_stimulus]$	2.723 (1.517)	0.073
$1[Lyon]$	0.146 (0.482)	0.762
Age	0.130 (0.090)	0.150
$Pseudo-R^2$	0.065	
$Prob > \chi^2$	0.401	
N	116	

Note. Dependent variable: player B from the stimuli rolled a die in the hidden action game (0/1). Pooled stimuli collected in Lyon ($1[Lyon]=1$) and in Paris ($1[Lyon]=0$). Each of the seven characteristics (Beautiful, Trustworthy, Intelligent, Self-confident, Masculine, Competent, Predictable) is measured on a 1 to 7 Likert scale (with higher values implying stronger agreement with the characteristic). $1[Female\_stimulus]=1$ if the player B seen in the stimulus is female, and $=0$ otherwise.

Table 4
Who is perceived as cooperative: OLS analysis.

	coeff (SE)	p
Intercept	-0.006 (0.187)	0.973
Beautiful	0.115 (0.021)	<0.001
Trustworthy	0.067 (0.025)	0.009
Intelligent	-0.043 (0.039)	0.271
Self-confident	-0.027 (0.016)	0.089
Masculine	0.035 (0.018)	0.051
Competent	-0.017 (0.045)	0.714
Predictable	-0.032 (0.027)	0.236
$1[Female\_stimulus]$	0.006 (0.052)	0.905
$1[Lyon]$	-0.038 (0.017)	0.026
Age	0.006 (0.002)	0.005
$1[Actual\_Roll]$	0.012 (0.014)	0.388
R^2	0.462	
$Prob > F$	<0.001	
N	116	

Note. Dependent variable: aggregate rate of prediction *Roll* assigned to a given player B by the raters from Nice. Pooled stimuli collected in Lyon ($1[Lyon]=1$) and in Paris ($1[Lyon]=0$). Each of the seven characteristics (Beautiful, Trustworthy, Intelligent, Self-confident, Masculine, Competent, Predictable) is measured on a 1 to 7 Likert scale (with higher values implying stronger agreement with the characteristic). $1[Female\_stimulus]=1$ if the player B seen in the stimulus is female, and $=0$ otherwise. $1[Actual\_Roll]=1$ if a player B actually rolled a die, and $=0$ otherwise.

To address this question, in Table 4 we present the results of a linear regression model in which the dependent variable is the stimulus-level rate of prediction *Roll*. The explanatory variables are the same as in the model in Table 3. We additionally include a dummy variable regarding the player B's actual decision ($1[Actual\_Roll]=1$ if the player B rolled, and $=0$ otherwise).[10] As we can see from Table 4, ratings of beauty and trustworthiness (previously shown to be orthogonal to the actual cooperativeness) condition the perceived cooperativeness. The strongest marginal effect is observed for the beauty ratings: ceteris paribus, a one-point increase in beauty leads to a 12 percentage point increase in the predicted rate of cooperativeness. In addition, the beliefs regarding cooperativeness (i.e., cooperativeness ratings) are not associated with the actual observed behavior from the earlier stage: the coefficient of the variable $1[Actual\_Roll]$ is close to zero and statistically insignificant.[11]

[10] The aggregate rate of prediction *Roll* assigned to a target player B in stimulus i is modeled as follows: $Rate\_of\_prediction\_Roll_i = \beta_0 + \beta_1 \times Beautiful_i + \beta_2 \times Trustworthy_i + \beta_3 \times Intelligent_i + \beta_4 + Self-confident_i + \beta_5 \times Masculine_i + \beta_6 \times Competent_i + \beta_7 \times Predictable_i + \beta_8 \times 1[Female\_stimulus_i] + \beta_9 \times 1[Lyon_i] + \beta_{10} \times Age_i + \beta_{11} \times 1[Actual\_Roll_i] + u_i$.

[11] The average rate of prediction *Roll* for cooperative (0.540) and non-cooperative player Bs (0.524) are not statistically different (t-test, $p=0.378$).

Table 5
Beauty bias and own beauty: logistic regression analysis.

	coeff (SE)	p	coeff (SE)	p
Intercept	-1.363 (0.593)	0.022	-1.551 (1.969)	0.431
Beautiful–self	-0.018 (0.118)	0.877	0.024 (0.440)	0.957
Beautiful–stimulus	0.311 (0.069)	<0.001	0.355 (0.448)	0.428
Beautiful–self × Beautiful–stimulus			-0.010 (0.099)	0.922
$Pseudo - R^2$	0.009		0.009	
$Prob > \chi^2$	<0.001		<0.001	

Note. Dependent variable: prediction *Roll* assigned to a given stimuli (issued from the Parisian sample) by a participant in Lyon. N = 1780 (89 individual-level clusters, 20 observations per cluster, cluster-robust SE). "Beautiful" is measured through an increasing scale ranging from 1 to 7.

Altogether, we document the existence of a systematic bias in judgments of cooperativeness. Judgments of cooperativeness are dissociated from actual behavior. However, raters condition their judgments of cooperation on factors that are not predictive of the actual behavior, with an important role played by beauty.[12] Finally, the evidence presented in Section 4.1 rules out confounding gender effects: while female player Bs is perceived as more beautiful, no such gender difference arises in terms of the rate of prediction $Roll$.

In the next part, we investigate the possible source of heterogeneity in the beauty bias: an interplay between one's beauty and the extent of conditioning one's judgments about other people's cooperativeness on their beauty.

4.3. Beauty bias: the role of the rater's beauty

Having documented the strong and factually uninformative beauty bias in the previous section, we now turn to investigating whether individuals that are themselves rated as more beautiful display more or less of such a bias.

Table 5 reports estimates from a logistic regression model relating the likelihood of making prediction $Roll$ to one's own beauty rating and the overall beauty rating of the player B seen in the stimulus (see the results on the left section of the table).[13] The model confirms the previously reported beauty bias: stimuli with higher beauty ratings have a higher likelihood of receiving a prediction of $Roll$. In contrast, the variation in the rater's beauty rating does not affect the likelihood of prediction $Roll$ ($p = 0.877$).

In a second specification (presented in the right section of the table), we include interaction terms between rater and stimulus beauty scores. We cannot reject the null hypothesis that there is no interaction between the two beauty ratings concerning their impact on the likelihood of prediction $Roll$ (the estimated coefficient is close to 0; $p = 0.922$).[14] Hence, beautiful decision-makers are neither more nor less inclined to believe that other beautiful individuals are cooperative.

5. Conclusions

In this paper, we study cooperativeness judgments and their precision for and by beautiful people. A sizable prior literature has documented a strong beauty bias with respect to the judgment of cooperativeness and trustworthiness of others. We add to this by documenting that the bias is a decision-making error: while beauty is a key predictor of cooperativeness ratings collected in our sample, it does not predict actual cooperativeness in an incentivized hidden-action game.

Additionally, we investigate the dependence of the raters' ratings and beliefs on their own beauty.

Beautiful individuals may judge beauty in others in different ways: giving it either more or less weight compared to those less beautiful. Our design elicits cooperativeness predictions by individuals that have been externally rated regarding their own beauty, thus allowing us to investigate whether and how the exhibited beauty bias changes among more or less beautiful individuals. We observe that own beauty is not associated with any difference in the beauty bias.

Overall, our results confirm the importance and robustness of the beauty bias when it comes to judging the cooperativeness of others. Overcoming the bias seems particularly hard as even beautiful individuals (who could have learned about its faultiness) fall prey to it. This has important consequences when it comes to business or labor interactions. It is known that most people attribute high values to seeing the face of others and that most people are keen to see photos of others before deciding to engage in cooperative activities with them (Eckel and Petrie, 2011; Ewing et al., 2015). In consequence, many job interviews are still face-to-face, as are most negotiations. For instance, more than 80% of business executives who responded to a recent Forbes survey preferred face-to-face meetings to virtual ones.[15] The main reason is that they feel that the former helps build stronger, more meaningful relationships

[12] In addition to the regression estimates, we observe a substantial correlation between the stimulus-level rate of prediction $Roll$ and the beauty rating ($\rho = 0.461$, $p < 0.001$). As shown in left-hand side column of Table 7 in Appendix C, this is the highest correlation coefficient between the rate of prediction $Roll$ and any of the seven individual characteristics of interest.

[13] Note that this does not include ratings by the individual themselves and thus avoiding any potential consistency biases. Both the beautiful-self and the beautiful-stimulus rating were made by raters from Stage 3.

[14] We note that including the interaction term preserves coefficient estimates, but (as could be expected) inflates the estimated standard errors due to multicollinearity. As a consequence, although none of the variable coefficients in the second specification happens to be significant, we strongly reject the null hypothesis of their joint insignificance.

[15] "Business meetings: The case for face-to-face" survey conducted by Forbes Insights in 2009.

and facilitates the reading of another person. In line with this argument, recent experimental research shows that observing rich social signals, e.g., pre-play interactions with face-to-face communication, helps build trust (Babutsidze et al., 2021) and predict cooperativeness (see, e.g., Zylbersztejn et al., 2020, 2021). Our study, in turn, shows how using observable information may backfire when social signals are scarce and restricted to another person's face. Cooperativeness judgments of faces are fast and present even when faces are not consciously seen (Todorov et al., 2009; Freeman et al., 2014; De Neys et al., 2017). The bias works swiftly, consistently, and unperceived. To avoid falling prey to it, businesses and individuals should consciously abstain from seeing pictures of potential partners, a suggestion that might be counter-intuitive for many. Overall, the results of our study call for rules and recommendations that will suppress this tendency and enable individuals to make less biased judgments.

Declaration of competing interest

I, Astrid Hopfensitz, declare that neither I nor any of my co-authors on this paper have any conflict of interest.

Data availability

Data will be made available on request.

Appendix A. Instructions for Stages 1 and 2

Note: Part 1 presents the hidden action game and is common for Stages 1 and 2. Part 2 presents the prediction task and is specific to Stages 2.

Part 1

You will now play a game with monetary stakes. The rules of the game are as follows.

The game is played by two players: player A and player B. Each player must choose between two possible actions. Player A chooses between actions "Left" and "Right". Player B chooses whether she wants a six-sided die to be rolled (action "Roll") or not (action "Don't roll").

Each player's payoff depends on the actions chosen by herself as well as the other player:

- if player A chooses "Left", then regardless of player Bs' choice:
  - player A's payoff is 5 EUR and player B's payoff is 5 EUR;
- if player A chooses "Right" and player B chooses "Don't roll":
  - player A's payoff is 0 EUR and player B's payoff is 14 EUR;
- if player A chooses "Right" and player B chooses "Roll":
  - if the number of on the die is between 1 and 5, then player A's payoff is 12 EUR and player B's payoff is 10 EUR;
  - if the number of on the die is 6, then player A's payoff is 0 EUR and player B's payoff is 10 EUR;

How the game proceeds

Stage 1. You are randomly assigned to your role – either player A or player B. A message on your computer screen will inform you about your role.
Stage 2. All participants are randomly and anonymously matched into pairs: if your are player A, then player A B is randomly selected to your complete pair; analogously, if your are player B, then player A A is randomly selected to complete your pair.
Stage 3. Player A chooses between "Left" and "Right" by clicking on a relevant button on his/her computer screen.
Stage 4. Player B chooses between "Don't roll" and "Roll" by clicking on a relevant button on his/her computer screen.

The outcome of the game for each pair of players is determined by the decisions made by both players. If the decisions in a pair of players are "Right" and "Roll", then the computer program rolls a die to determine players' payoffs.

At the end of the experiment, players are only informed about their personal payoffs, and not about the payoffs of or the decisions made by other players, or about the outcome of the die roll.

Part 2

In this part, you will be asked to predict the decisions other people previously made in another experiment.

In that previous experiment, participants received the same instructions and played the same game as you did in Part 1 of the present experiment. Your predictions in this part will be related to the decisions made by the participants acting as player Bs in the previous experiment.

Your role

This part of the experiment consists of **multiple rounds**. In each round, you will be asked to **make a prediction** about a player B's behavior in the previous experiment.

At the beginning of each round, you will see a picture. Each **picture presents a person in the role of player B from the previous experiment** (as described above). The picture has been taken privately and independently of the previous experiment.

After seeing a picture, you will be asked to predict if the player B from that picture decided to roll a die in the previous experiment. Your gain will depend on the accuracy of your prediction. A **correct prediction is worth 100 points**, while an **incorrect one is worth 20 points** (with **1 point = 0.10 EUR**).

Please note that overall, exactly **50% of the player Bs you will see in the pictures decided to roll a die** in the previous experiment, while the **other 50% did not**.

At the end of the experiment, **one round will be chosen at random** by the computer, and the prediction made in that round may determine your payoff in this part of the experiment.

Your payoff

Your final payoff in this part of the experiment will be determined as follows.

Before starting the first round, you will be also asked to state an amount (denoted M) between 21 and 100 points. This **amount should make you indifferent** between receiving M for sure (and without the need to predict player B's choice) and being paid based on the randomly drawn prediction about player A B's choice which, depending on your accuracy, may earn you either 100 points (for a correct prediction) or 20 points (for an incorrect prediction).

Once the last round completed, **one round will be chosen at random** by the computer software (each round being equally probable). The computer will also randomly generate a number between 20 and 99 (denoted N), all values within this range being equally probable.

Then, the procedure runs as follows:

- **If** $M \leq N$ (that is, if the amount you have chosen for the randomly selected task is smaller than or equal to the random number N generated by the computer), then you will receive a **certain payoff of** $N \times 0.10$ **EUR**. This means that your payoff will not depend on your prediction in the randomly selected round.
- **If** $M > N$ (that is, if the amount you have chosen for the randomly selected task is greater than the random number N generated by the computer), then your payoff will depend on your prediction (i.e., either the prediction made in a randomly selected round): **you will earn 100 points (10 EUR) for a correct prediction** or **20 points (2 EUR) for an incorrect prediction**.

Here are some examples illustrating this procedure:

Example 1. If the amount M you chose for the task randomly selected by the computer equals 31, and the randomly generated number N equals 24, then your payoff will depend on the prediction made in the task selected by the computer.

Example 2. If the amount M you chose for the task randomly selected by the computer equals 45, and the randomly generated number N equals 76, then your gain will be equal to 7.60 EUR (76×0.10 EUR) and your payoff will not depend on the prediction made in the round selected by the computer.

Example 3. If the amount M you chose for the task randomly selected by the computer equals 81, and the randomly generated number N equals 81, then your gain will be equal to 8.10 EUR (81×0.10 EUR) and your payoff will not depend on the prediction made in the round selected by the computer.

Appendix B. Instructions for Experiment 3

In this part of the experiment, you will see a series of pictures of people.

You will be asked to predict the decisions those people previously made in another experiment (the details of which are described below). Your final gain will depend on the accuracy of your predictions.

The previous experiment

In an experimental economics laboratory, participants were seated in front of computers and took part in an interactive game with monetary stakes. No communication between participants was allowed. Decisions remained strictly anonymous during and after the experiment.

Rules of the game

The game is played by two players: player A and player B. Each player must choose between two possible actions. Player A chooses between actions “Left” and “Right”. Player B chooses whether she wants a six-sided die to be rolled (action “Roll”) or not (action “Don't roll”).

Each players' payoff depends on the actions chosen by herself as well as the other player:

- if player A chooses "Left", then regardless of player Bs' choice:
  - player A's payoff is 5 EUR and player B's payoff is 5 EUR;
- if player A chooses "Right" and player B chooses "Don't roll":
  - player A's payoff is 0 EUR and player B's payoff is 14 EUR;
- if player A chooses "Right" and player B chooses "Roll":
  - if the number of on the die is between 1 and 5, then player A's payoff is 12 EUR and player B's payoff is 10 EUR;
  - if the number of on the die is 6, then player A's payoff is 0 EUR and player B's payoff is 10 EUR;

How the game proceeds

Stage 1. You are randomly assigned to your role – either player A or player B. A message on your computer screen will inform you about your role.
Stage 2. All participants are randomly and anonymously matched into pairs: if your are player A, then player A B is randomly selected to your complete pair; analogously, if your are player B, then player A A is randomly selected to complete your pair.
Stage 3. Player A chooses between "Left" and "Right" by clicking on a relevant button on his/her computer screen.
Stage 4. Player B chooses between "Don't roll" and "Roll" by clicking on a relevant button on his/her computer screen.

The outcome of the game for each pair of players is determined by the decisions made by both players. If the decisions in a pair of players are "Right" and "Roll", then the computer program rolls a die to determine players' payoffs.

At the end of the experiment, players are only informed about their personal payoffs, and not about the payoffs of or the decisions made by other players, or about the outcome of the die roll.

Your role

This part of the experiment consists of **multiple rounds**. In each round, you will be asked to **make a prediction** about player A B's behavior in the previous experiment.

At the beginning of each round, you will see a picture. Each **picture presents a person in the role of player B from the previous experiment** (as described above). The picture has been taken privately and independently of the previous experiment.

After seeing a picture, you will be asked to predict if the player B from that picture decided to roll a die in the previous experiment. Your gain will depend on the accuracy of your prediction. A **correct prediction is worth 100 points**, while an **incorrect one is worth 20 points** (with **1 point = 0.10 EUR**).

Please note that overall, exactly **50% of the player Bs you will see in the pictures decided to roll a die** in the previous experiment, while the **other 50% did not**.

At the end of the experiment, **one round will be chosen at random** by the computer, and the prediction made in that round may determine your payoff in this part of the experiment.

Appendix C. Complementary evidence on effects sizes

Table 6
Individual ratings and decisions in the hidden action game: effect sizes.

Average score:	Cohen's d
Rate of prediction *Roll*	-0.164
Beautiful	-0.143
Trustworthy	-0.060
Intelligent	-0.163
Self-confident	0.040
Masculine	-0.170
Competent	-0.147
Predictable	0.026

Note. Pooled stimuli (N = 116) collected in Lyon and in Paris. Group variable: individual decision to *Roll* (N = 60) vs. *Don't roll* (N = 56) in the hidden action game. Individual rate of prediction *Roll* varies between 0 and 1. Each of the seven characteristics (Beautiful, Trustworthy, Intelligent, Self-confident, Masculine, Competent, Predictable) is measured on a 1 to 7 Likert scale (with higher values implying stronger agreement with the characteristic). ***/** indicate statistical significance at the 1%/5% level.

Table 7
Spearman's correlation coefficients across rated characteristics.

Average score:	Rate of pred. *Roll*	Beautiful	Trustworthy	Intelligent	Self-conf.	Masculine	Competent	Predictable
Rate of pred. *Roll*	1.000							
Beautiful	0.461***	1.000						
Trustworthy	0.430***	0.564**	1.000					
Intelligent	0.205**	0.417**	0.502**	1.000				
Self-confident	0.108	0.266**	-0.241**	-0.128	1.000			
Masculine	-0.112	-0.684**	-0.433***	-0.131	0.003	1.000		
Competent	0.250***	0.375**	0.520**	0.848**	-0.006	-0.043	1.000	
Predictable	0.171	0.227	0.560**	0.159	-0.508**	-0.301**	0.068	1.000

Note. Pooled stimuli ($N = 116$) collected in Lyon and in Paris. Individual rate of prediction *Roll* varies between 0 and 1. Each of the seven characteristics (Beautiful, Trustworthy, Intelligent, Self-confident, Masculine, Competent, Predictable) is measured on a 1 to 7 Likert scale (with higher values implying stronger agreement with the characteristic). ***/** indicate statistical significance at the 1%/5% level.

References

Agthe, M., Spörrle, M., Maner, J.K., 2011. Does being attractive always help? Positive and negative effects of attractiveness on social decision making. Pers. Soc. Psychol. Bull. 37, 1042–1054.

Altman, H.J.R., Altman, M., Torgler, B., Whyte, S., 2021. Beauty, Preferences and Choice Exemplified in the Sports Market. Behavioural Sports Economics: A Research Companion, vol. 201.

Babutsidze, Z., Hanaki, N., Zylbersztejn, A., 2021. Nonverbal content and trust: an experiment on digital communication. Econ. Inq. 59, 1517–1532.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., 2001. The "Reading the Mind in the Eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J. Child Psychol. Psychiatry Allied Discipl. 42, 241–251.

Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9, 226–232.

Belot, M., Bhaskar, V., Van De Ven, J., 2012. Beauty and the sources of discrimination. J. Hum. Resour. 47, 851–872.

Bokek-Cohen, Y., Davidovich, N., 2011. The beauty premium in the academic world: a cross-cultural perspective. Kentucky J. Excell. Coll. Teach. Learn. 7, 36–44.

Bovet, J., 2019. Evolutionary theories and men's preferences for women's waist-to-hip ratio: which hypotheses remain? A systematic review. Front. Psychol. 10, 1221.

Budesheim, T.L., DePaola, S.J., 1994. Beauty or the beast? The effects of appearance, personality, and issue information on evaluations of political candidates. Pers. Soc. Psychol. Bull. 20, 339–348.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74, 1579–1601.

Chen, D.L., Hopfensitz, A., van Leeuwen, B., van de Ven, J., 2019. The Strategic Display of Emotions. CentER Discussion Paper Series.

Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B., Wu, C.-H., 1995. 'Their ideas of beauty are, on the whole, the same as ours': consistency and variability in the cross-cultural perception of female physical attractiveness. J. Pers. Soc. Psychol. 68, 261–279.

Dasgupta, N., Rivera, L.M., 2008. When social context matters: the influence of long–term contact and short–term exposure to admired outgroup members on implicit attitudes and behavioral intentions. Social Cogn. 26, 112–123.

De Neys, W., Hopfensitz, A., Bonnefon, J.-F., 2017. Split-second trustworthiness detection from faces in an economic game. Exp. Psychol. 64, 231–239.

Deryugina, T., Shurchkov, O., 2015. Does beauty matter in undergraduate education? Econ. Inq. 53, 940–961.

Devine, P.G., 1989. Stereotypes and prejudice: their automatic and controlled components. J. Pers. Soc. Psychol. 56, 5–18.

Dion, K., Berscheid, E., Walster, E., 1972. What is beautiful is good. J. Pers. Soc. Psychol. 24, 285–290.

Doorley, K., Sierminska, E., 2015. Myth or fact? The beauty premium across the wage distribution in Germany. Econ. Lett. 129, 29–34.

Eckel, C., Petrie, R., 2011. Face value. Am. Econ. Rev. 101, 1497–1513.

Ewing, L., Caulfield, F., Read, A., Rhodes, G., 2015. Perceived trustworthiness of faces drives trust behaviour in children. Dev. Sci. 18, 327–334.

Feingold, A., 1992. Good-looking people are not what we think. Psychol. Bull. 111, 304–341.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.

Frederick, S., 2005. Cognitive reflection and decision making. J. Econ. Perspect. 19, 25–42.

Freeman, J.B., Stolier, R.M., Ingbretsen, Z.A., Hehman, E.A., 2014. Amygdala responsivity to high-level social information from unseen faces. J. Neurosci. 34, 10573–10581.

Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. Q. J. Econ. 112, 631–645.

Gneezy, U., Rustichini, A., Vostroknutov, A., 2010. Experience and insight in the Race game. J. Econ. Behav. Organ. 75, 144–155.

Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1, 114–125.

Hamermesh, D.S., 2013. Beauty Pays: Why Attractive People Are More Successful. Princeton University Press.

Hamermesh, D.S., Biddle, J.E., 1994. Beauty and the labor market. Am. Econ. Rev. 84, 1174–1194.

Kalick, S.M., Zebrowitz, L.A., Langlois, J.H., Johnson, R.M., 1998. Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. Psychol. Sci. 9, 8–13.

Legault, L., Gutsell, J.N., Inzlicht, M., 2011. Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice. Psychol. Sci. 22, 1472–1477.

Mann, N.H., Kawakami, K., 2012. The long, steep path to equality: progressing on egalitarian goals. J. Exp. Psychol. Gen. 141, 187–197.

Mitchem, 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. Evol. Hum. Behav. 36, 240–247.

Ramos, M.R., Barreto, M., Ellemers, N., Moya, M., Ferreira, L., Calanchini, J., 2016. Exposure to sexism can decrease implicit gender stereotype bias. Eur. J. Soc. Psychol. 46, 455–466.

Rhodes, G., Zebrowitz, L.A., Clark, A., Kalick, S.M., Hightower, A., McKay, R., 2001. Do facial averageness and symmetry signal health? Evol. Hum. Behav. 22, 31–46.

Ruffle, B.J., Sherman, A., Shtudiner, Z., 2022. Gender and beauty price discrimination in produce markets. J. Behav. Exp. Econ. 97, 101825.

Scholz, J.K., Sicinski, K., 2015. Facial attractiveness and lifetime earnings: evidence from a cohort study. Rev. Econ. Stat. 97, 14–28.

Snyder, M., Tanke, E.D., Berscheid, E., 1977. Social perception and interpersonal behavior: on the self-fulfilling nature of social stereotypes. J. Pers. Soc. Psychol. 35, 656–666.

Solnick, S.J., Schweitzer, M.E., 1999. The influence of physical attractiveness and gender on ultimatum game decisions. Organ. Behav. Hum. Decis. Process. 79, 199–215.

Symons, D., 1979. The Evolution of Human Sexuality. Oxford University Press.

Teng, F., Wang, X., Zhang, Y., Lei, Q., Xiang, F., Yuan, S., 2022. Mirror, mirror on the wall, I deserve more than all: perceived attractiveness and self-interested behavior. Evol. Hum. Behav. 43, 536–547.

Thornhill, R., Gangestad, S.W., 1999. Facial attractiveness. Trends Cogn. Sci. 3, 452–460.
Todorov, A., Pakrashi, M., Oosterhof, N.N., 2009. Evaluating faces on trustworthiness after minimal time exposure. Social Cogn. 27, 813–833.
Van Leeuwen, B., Noussair, C.N., Offerman, T., Suetens, S., Van Veelen, M., Van De Ven, J., 2018. Predictably angry—facial cues provide a credible signal of destructive behavior. Manag. Sci. 64, 3352–3364.
Vogt, S., Efferson, C., Fehr, E., 2013. Can we see inside? Predicting strategic behavior given limited information. Evol. Hum. Behav. 34, 258–264.
Weeden, J., Sabini, J., 2005. Physical attractiveness and health in western societies: a review. Psychol. Bull. 131, 635–653.
Wilson, R.K., Eckel, C.C., 2006. Judging a book by its cover: beauty and expectations in the trust game. Polit. Res. Q. 59, 189–202.
Zebrowitz, L.A., Hall, J.A., Murphy, N.A., Rhodes, G., 2002. Looking smart and looking good: facial cues to intelligence and their origins. Pers. Soc. Psychol. Bull. 28, 238–249.
Zylbersztejn, A., Babutsidze, Z., Hanaki, N., 2020. Preferences for observable information in a strategic setting: an experiment. J. Econ. Behav. Organ. 170, 268–285.
Zylbersztejn, A., Babutsidze, Z., Hanaki, N., 2021. Predicting trustworthiness across cultures: an experiment. Front. Psychol. 12, 727550.