Title:

Intelligence can be detected but is not found attractive in videos and live interactions

Driebe, Julie C.*[a,1]

Sidari, Morgan J.*[a,2]

Dufner, Michael[3]

von der Heiden, Juliane M.[4]

Bürkner, Paul C.[5]

Penke, Lars[b, 1,6]

Zietsch, Brendan P.[b,2]

Arslan, Ruben C.[b,7]

[a] share first authorship, [b] share last authorship

* corresponding author for study 1: julie.driebe@uni-goettingen.de, for study 2: morgan.sidari@uqconnect.edu.au

[1] Biological Personality Psychology, University of Goettingen, Germany

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

[3] Personality Psychology and Psychological Assessment, Witten/Herdecke University, Germany

[4] Personality Psychology and Psychological Assessment, University of Mainz, Germany

[5] Computer Science, Aalto University, Finnland

[6] Leibniz ScienceCampus Primate Cognition, Goettingen, Germany

[7] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Abstract

Humans' extraordinary intelligence seems to extend beyond the needs for survival. One theory to explain this surplus intelligence is that it evolved via sexual selection as a fitness indicator to advertise genetic quality to prospective mates. Consistent with this idea, self-reported mate preferences suggest intelligence is valued across cultures. Yet, as the validity of these self-reports has been questioned, it remains unclear whether objectively assessed intelligence is indeed attractive. We analysed data from two studies to test this key premise of the sexual selection theory of intelligence. In Study 1, 88 target men had their intelligence measured and based on short video clips were rated on intelligence, funniness, physical attractiveness and mate appeal by 179 women. In Study 2 (N = 729), participants took part in 2 to 5 speed-dating sessions in which their intelligence was measured and they rated each other's intelligence, funniness, and mate appeal. Measured intelligence did not predict increased mate appeal in either study, whereas perceived intelligence and funniness did. More intelligent people were perceived as more intelligent, but not as funnier. Results suggest that intelligence is not important for initial attraction, which raises doubts concerning the sexual selection theory of intelligence.

Humans' extraordinary intelligence is an important aspect that distinguishes us from all other animals; however, the evolutionary forces that gave rise to this peculiar feature are not well understood. In our evolutionary past, it is likely that intelligence assisted our ancestors in solving adaptive problems, such as evading predators, navigating social hierarchies, and crafting complex tools. However, our intelligence seems to go far beyond what is required for mere survival, as it enables us to compose music, create art and literature, and to engage in humorous wordplay. Such activities do not have clear survival benefits, and indeed the human brain's energy demands are enormous relative to the other organs of the human body and the brains of other animals[1]. One theory is that our surplus of intelligence has emerged through sexual selection[2,3]. Sexual selection refers to fitness differences between individuals that results from the number and/or quality of mates, and can result from differential attractiveness to the opposite-sex (*intersexual selection*) or ability to outcompete same-sex rivals (*intrasexual selection*)[4].

Miller[2] proposed that our surplus intelligence emerged through intersexual selection. As 84% of human genes are expressed in the brain, developing a healthy, optimally functioning brain requires an individual to be relatively free from harmful mutations[2,3,5,6]. For this reason, intelligence may be a revealing signal of genetic quality to potential romantic partners. Traits of this kind are referred to as fitness indicators, and can develop through the tandem influence of natural and sexual selection, solving both an adaptive problem in the evolutionary environment as well as advertising (to prospective mates) a superior capability to do so. Alternatively, an indicator can develop solely for the purpose of displaying fitness without any clear survival benefit. While general intelligence is clearly relevant to survival, its more elaborate manifestations, such as language complexity and humour, may be examples of fitness indicators that exist primarily for the purpose of displaying fitness[2,3].

If human intelligence evolved via romantic and sexual choices across multitudes of generations, this legacy should be reflected in our romantic and sexual preferences today[2,3,7]. Previous research has found that intelligence is one of the most desirable traits in a hypothetical partner[8] and this has been replicated cross-culturally[9]; however, speed-dating experiments have shown that these hypothetical preferences do not always correspond to actual initial mate choices made in live interactions[10]. This discrepancy could arise because people cannot accurately report their true preferences, or because some traits, such as intelligence, become important only at later stages of relationship formation[11].

To further complicate matters, research examining whether intelligence can be detected accurately has yielded mixed results, particularly in live contexts[12,13]. If attraction to intelligence evolved to select partners of high genetic quality, the attractiveness of intelligence should be seen in the choices individuals make in actual interactions. Further, it should not just extend to partners who we perceive to be intelligent - it should extend to partners with higher actual intelligence.

For a trait to be intersexually selected, either the trait itself or a correlated trait must be detectable by members of the opposite sex[2,3]. This detection can be conscious or subconscious[14]. In the case of intelligence, a large body of literature referred to as *thin-slicing* has sought to determine whether measured intelligence is detectable. Thin-slicing is a paradigm in which one individual is first objectively measured on a given trait. This individual (the actor) is then rated on the same trait by another individual (the observer) after a brief exposure such as a photograph, recording or interaction[12]. As with many cognitive and personality traits, thin-slicing research on intelligence suggests that it can be estimated by the observer with some degree of accuracy[12,15-17]. Further, moderate accuracy can be achieved with exposure to video clips as brief as five seconds[18]. Previous work suggests that estimating intelligence after a live interaction is difficult, so that video judges outperformed live

interaction partners, although the study was not adequately powered to distinguish negligible from low accuracy[16]. Thus, there is a low to moderate overlap between a target person's actual level of intelligence and observer-rated intelligence, which also indicates that the actual effects of intelligence on mate appeal might be very different from what people report about their mate preferences. In order to test whether intelligence is indeed predictive of mate appeal, research should not rely on self-reported partner preferences, but rather have participants rate the mate appeal of individuals who also had their intelligence tested objectively. We are aware of only Prokosch et al. (2009)[19] who conducted such a study. The researchers investigated the influence of both objectively measured and perceived intelligence on mate appeal. Female participants (204 in total) watched a series of videos of men performing verbal and physical tasks, such as responding to a thought-provoking question, reading newspaper headlines aloud and throwing a Frisbee. Prokosch et al.[19] found that both measured and rated intelligence predicted mate appeal, although measured intelligence had a small effect. The study had several limitations. Intelligence was assessed using only a vocabulary subtest, only 15 men were recruited as targets, women rated only 5 men each, and the same women rated intelligence and mate appeal, potentially inducing transfer effects. Thus, the study did not contain a comprehensive assessment of intelligence, it had low statistical power and common method bias.

A final topic that needs to be discussed in this context is the role of humour. People in courtship situations typically do not solve any cognitive tasks, which would make it quite easy to directly infer intelligence. Miller's sexual selection theory of intelligence suggests that potential partners instead often rely on a cue that is indicative of high intelligence, namely humour[2]. Humour, or funniness, does not confer any clear direct benefits to potential partners or offspring but is rated as one of the most attractive traits for a prospective partner[20-22]. Its attractiveness in the absence of clear survival benefits is consistent with humour

displaying underlying traits, such as intelligence, that indicate genetic quality. Previous research has connected measured intelligence to humour production in tasks such as producing captions for cartoons or wordplay, but these tasks may have limited resemblance to the interpersonal humour most relevant for courtship[23,24]. The role of interpersonal humour in directly or indirectly indicating intelligence therefore remains an open question.

The Current Study

In our two studies, we investigated the accuracy of intelligence judgements based on short sequences of behaviour[25] and the impact of intelligence on mate appeal, as well as on perceptions of funniness. In study 1, we used highly controlled conditions (i.e. short video sequences of participants), comprehensive intelligence measures, a large target sample size, and a repeated measures design that assessed women's judgements multiple times as the information on targets' intelligence increases. The purpose of this repeated measures design is that by gradually presenting different cues with increasing intelligence information above cues on only physical attractiveness, we can isolate the effect of intelligence on mate appeal (see Figure 6). In addition, different samples of women rated either intelligence, funniness, physical attractiveness or mate appeal to reduce transfer effects and shared response tendencies. These changes allow us to determine how mate appeal of targets changes with more information about their intelligence and funniness while, importantly, allowing us to control for potential halo effects. According to Miller's hypothesis, the preference for intelligence should be stronger among female, as compared to male raters[9,20]. Hence, testing women's preferences is a powerful test of the hypothesis.

In study 2, we adopted a more ecologically valid speed-dating methodology whereby participants' verbal intelligence was measured and they provided ratings on each other's intelligence, funniness and mate appeal after a 3-minute meeting. Parts of study 1 were preregistered (https://osf.io/rs3tg/); however, during the course of the project we realised that

some predictions were insufficient and we opted for more appropriate analyses. For transparency, we have provided a table in our appendix (S1) which highlights the deviations from our preregistration and details their respective rationales.

Study 1. For intelligence to play a focal role in human mate choice, it needs to be perceived somewhat accurately. First, we predict that women's intelligence ratings for male targets, based on short sequences of behaviour (e.g. reading newspaper headlines aloud[14]), will be positively correlated with targets' psychometrically measured intelligence.

Second, we investigate the influence of funniness, a proposed more perceivable display of intelligence, on sexual mate appeal. We hypothesise that perceived funniness is associated with measured intelligence and that men's perceived funniness will predict their rated sexual mate appeal above and beyond the effect of their intelligence. Further, we hypothesise that perceived intelligence predicts rated sexual mate appeal.

Third, in line with Miller's[2] hypothesis, we predict that men's measured intelligence will be significantly positively correlated with women's ratings of men's sexual mate appeal. And fourth, we predict that the increase in men's mate appeal after adding additional cues related to intelligence (i.e. reading newspaper headlines aloud; making experimenter laugh) will depend on men's intelligence, such that the sexual mate appeal increase will be greater for more intelligent men.

Study 2. For study 2, the hypotheses follow a similar rationale. First, we predict that psychometrically measured intelligence will be positively correlated with speed-dating partners' perception of intelligence. Second, we predict that measured intelligence will be positively correlated with speed-dating partners' ratings of mate appeal and funniness. Third, we predict speed-dating partners' ratings of intelligence and ratings of funniness will be positively correlated with their ratings of mate appeal for the same target.

Results

Study 1

Target's intelligence level. Using eight intelligence subtests, we assessed our targets' measured intelligence (see Table S2). Results of cognitive ability tests are substantially intercorrelated, yielding a latent, general factor of intelligence, referred to as the *g factor*[26]. We conducted a principal component analysis and found that the first unrotated factor, the *g* factor, explained 37% of variance. This factor served as the criterion measure of the target's measured intelligence adopted in study 1.

Accuracy of intelligence perception. To investigate the accuracy of intelligence perception, we first correlated targets' *g* factor with an aggregated value of perceived intelligence using a Pearson product-moment correlation, $r = .34$, ($p < .001$; 95% CI [.14; .51], Figure 1A). Aggregated perceiver values are commonly used in accuracy research; however, aggregates tend to lead to inflated accuracy estimates[27] and should be interpreted with caution. Therefore, we also used disaggregated ratings to determine the accuracy of individual women's judgments of intelligence ($\beta = .22$, $p < .001$, 95% CI [.07; .28]) in a structural equation model with standard errors clustered by target, modelling *g* as a hierarchical latent variable to correct for measurement error (see S3A). The results from both methods support our first prediction, suggesting that women are able to perceive intelligence with some degree of accuracy based on our three cues (cue 4: videos of men reading newspaper headlines aloud, cue 5: performing a pantomime task and cue 6: trying to make the experimenter laugh).

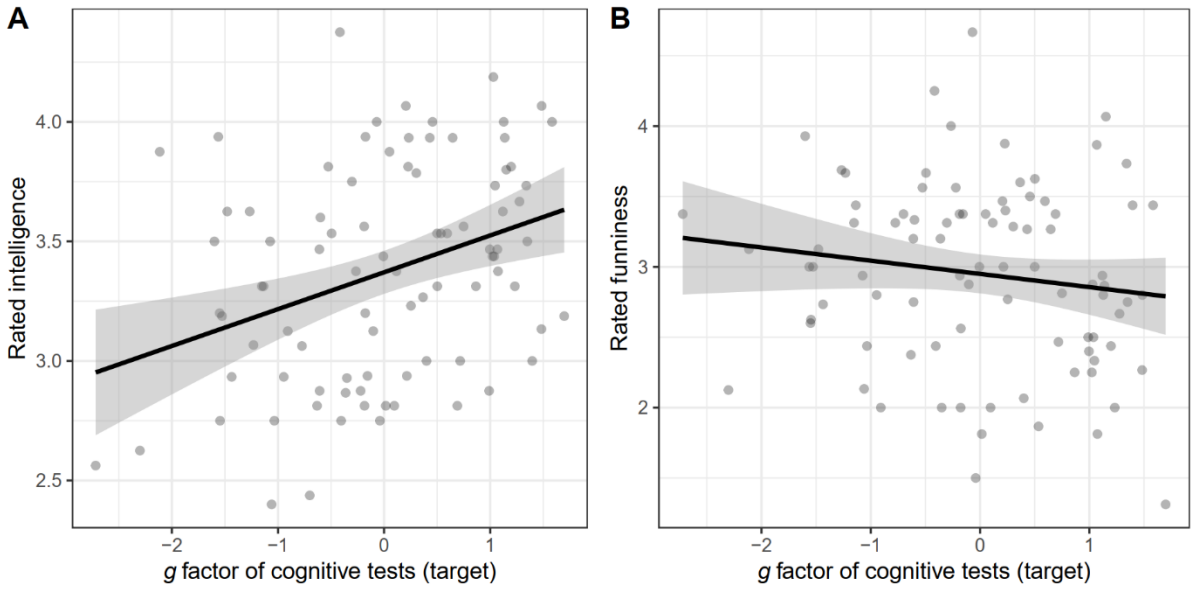


Figure 1. Aggregated perceiver accuracy for intelligence as measured by the *g* factor. The shaded area in grey reflects the 95% HDI.

Ratings of mate appeal. Women rated men's mate appeal operationalised as men's attractiveness as a short-term mate and long-term mate; however, we found that these ratings were highly correlated (r = .92). Therefore, all results are reported based on short-term mate attractiveness (henceforth referred to as sexual mate appeal); results for long-term mate attractiveness can be found in our supplement (see S3B).

Preference for funniness and perceived intelligence. If funniness is a display of intelligence, we would expect a relationship between men's measured intelligence and women's perception of men's funniness. Women's perception of men's funniness was associated with their perception of men's intelligence (b = .30, p > .001, 95% CI [.24; .36]). But contrary to expectations, measured intelligence was not associated with perceived funniness (r = -.14, p = .18, 95% CI [-.34; .07], Figure1B).

Further, we investigated whether funniness influences men's sexual mate appeal incremental to measured intelligence (Table 1). More intelligent men were rated to have a slightly lower sexual mate appeal (b = -.14, p = .03, 95% CI [-.26; -.01]), contrary to

expectations. However, men who were perceived to be more funny had a higher sexual mate appeal (b = .35, p < .001, 95% CI [.26; .45]). These results do not support the notion that funniness is a display of intelligence. We found that men who were perceived to be more intelligent also had a higher sexual mate appeal (b = .17, p = .002, 95% CI [.06, .29]) (Table S7).

Table 1: *LM coefficients for associations between measured intelligence, humour and sexual mate appeal*

	sexual mate appeal		
Term	Estimate	*p*	95 % CI
Intercept	.69	<.001	[.35; 1.03]
g factor	-.14	.03	[-.26; -.01]
Funniness	.35	<.001	[.26; .45]
Physical attractiveness	.24	<.001	[.17; .30]

Note. 88 Targets were rated by n = 30 women rating men's sexual mate appeal, n = 16 women rating men's funniness and n = 19 women rating men's physical attractiveness.

Preference for more intelligent men. Contrary to our prediction that women would prefer more intelligent men, we found that more intelligent men were rated to have a slightly lower sexual mate appeal (*g* factor: b = -.07, 95% HDI [- .11, -.03]). Men's physical attractiveness was the main predictor of sexual mate appeal (b = 1.15, 95% HDI [1.05; 1.24]) (see Table 2). These findings do not support our second prediction, suggesting that women did not find intelligent men more appealing.

Adding initial intelligence cues. We predicted that more intelligent men's sexual mate appeal would increase more than it would for less intelligent men when shifting from only physical attractiveness information being available (cue 1-3; various physical and vocal attractiveness cues) to provision of additional cues related to men's intelligence (cue 4; reading newspaper headlines, which has been a task strongly related to accurate intelligence perception in previous research[12]). As can be seen in Figure 2, after cue 4 was presented, the increase in men's sexual mate appeal ratings did not depend on their intelligence (*g* factor x cue 4: $b = .01$, 95% HDI [-.02; .04]). This finding does not support our prediction, in that cues of intelligence did not uniquely contribute to sexual mate appeal ratings. Additionally, we predicted that further adding information on men's funniness (cue 6; make experimenter laugh) would provide a greater increase in sexual mate appeal for more intelligent men. Cue 5 (pantomime) was not presented in order to reduce test fatigue (see Methods). Contrary to our predictions, we found that the increase in men's sexual mate appeal did not depend on their intelligence (*g* factor x cue 6: $b = .02$, 95% HDI [-.02; .04]). Taken together with the previous finding, this casts further doubt on the notion that intelligence is attractive in men.

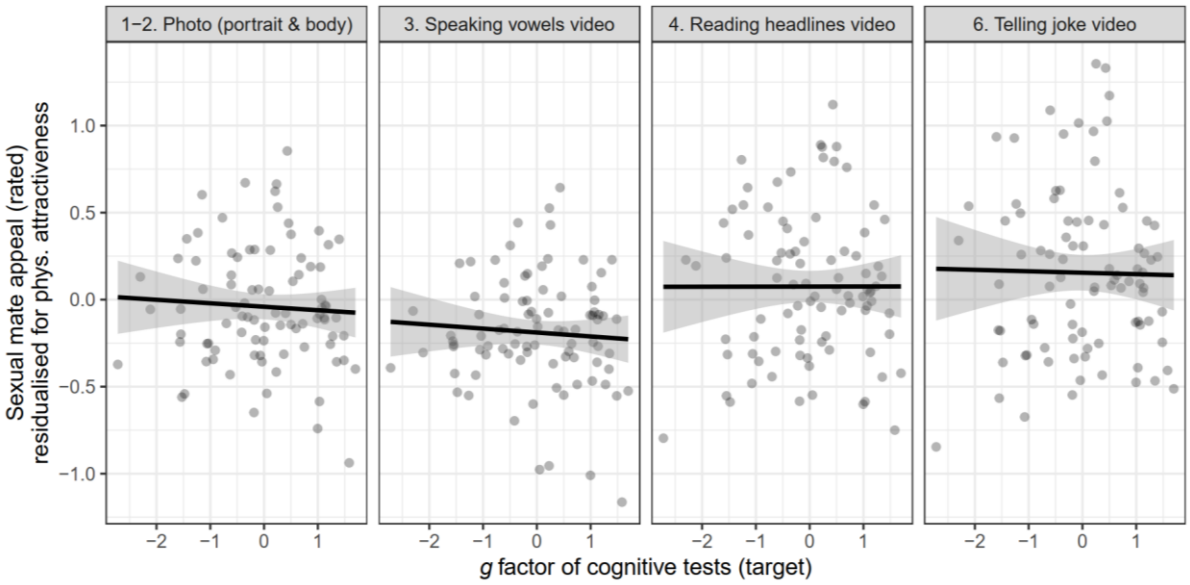


Figure 2. The aggregated sexual mate appeal ratings made after seeing each cue (or set of cues) was adjusted for physical attractiveness. The points shown in this plot show sexual mate appeal residualised for physical attractiveness. The shaded area in grey reflect the 95% HDI. The plot shows the slope of a linear regression predicting sexual mate appeal from the measured *g* factor. Intelligent men were not rated more favourably, even after intelligence-relevant information became available.

Table 2: *Associations between sexual mate appeal and measured intelligence in sequential cue presentation*

	sexual mate appeal	
Term	Estimate	95 % HDI
Intercept	.44	[0.09; 0.80]
Cue 1&2	-.30	[-.39; -.21]
Cue 4	.16	[.09; .23]
Cue 6	.36	[.25; .48]
physical attractiveness	1.15	[1.07; 1.22]
g factor	-.07	[-.10; -.03]

Cue 1&2 * physical attractiveness	-.10	[-.13; -.07]
Cue 4 * physical attractiveness	.07	[.04; .10]
Cue 6 * physical attractiveness	.14	[.11; .18]
Cue 1 & 2 * *g* Factor	-.01	[-.03; .02]
Cue 4 * *g* Factor	.01	[-.02; .04]
Cue 6 * *g* Factor	.02	[-.01; .05]

Note. Estimates and highest density intervals (HDI) from a Bayesian mixed effects location-scale model. Here, we show only the relevant non-varying effects on the mean, see Appendix S3B/online supportive materials for further control variables, varying effects and effects on scale. The reference category of the cue variable was set to the 'Vowels' video (cue 3), so that the interaction between cue 4 and measured intelligence captures the change in association at the point at which intelligence becomes task-relevant.

Additionally presented cues and attractiveness. Though the previous two results showed that change in sexual mate appeal with additional cues did not depend on men's intelligence, it should be noted that men's rated sexual mate appeal increased after cue 4 was presented (cue 4: b = .16, 95% HDI [.07; .24]) and further after cue 6 was presented (cue 6: b = .36, 95% HDI [.23; .50]). This raises the question of what other factor(s) involved in sexual mate appeal judgments were revealed in these later cues. We found that the increase in sexual mate appeal with additional stimuli was greater for more physically attractive men, with their ratings improving when after the presentation of cue 4 (cue 4 x physical attractiveness: b = .07, 95% HDI [.04; .11]) and the presentation of cue 6 (cue 6 x physical attractiveness: b = .14, 95% HDI [.10; .19]). Therefore, more physically attractive men did not only have a higher mate appeal, but they also benefited more from the later cues than did less physically attractive men.

Study 2

As predicted, more intelligent people were perceived to be more intelligent by their interaction partner, suggesting that intelligence is detectable in short live interactions (γ = 0.08, 95% CI [.03; .13], p = .002*)*. After aggregating ratings across raters, the correlation was r = .12 (Figure 3). However, contrary to predictions, more intelligent people were not more likely to be rated as funnier by their partners (γ = -0.01, 95% CI [-.06; .05], p = .841). We found no evidence that the associations between intelligence and perceptions differed by sex (ps > .91).

As predicted, men perceived to be more intelligent or funnier were also rated as having a higher mate appeal by their interaction partners. However, measured intelligence did not predict rated mate appeal (Table 3, Figure 4). We found no evidence that the associations with mate appeal differed by sex (ps > .18). Full results including random effects and moderation by sex can be found in the supplementary material F. Additionally, this pattern of results remained when controlling for both facial and bodily attractiveness, though some relationships between rated variables were attenuated. These results can be found in the supplementary material G.

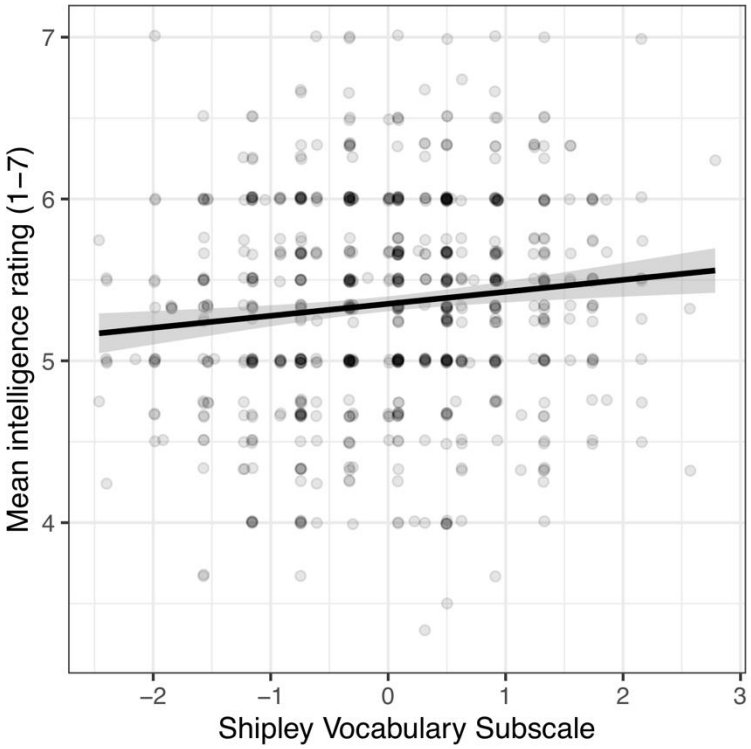


Figure 3. Association between intelligence, as measured by the Shipley Institute of Living Scale (Vocabulary Subscale), and rated intelligence, after aggregating across raters. Varying opacity of the dots is caused by overlap of multiple participants.

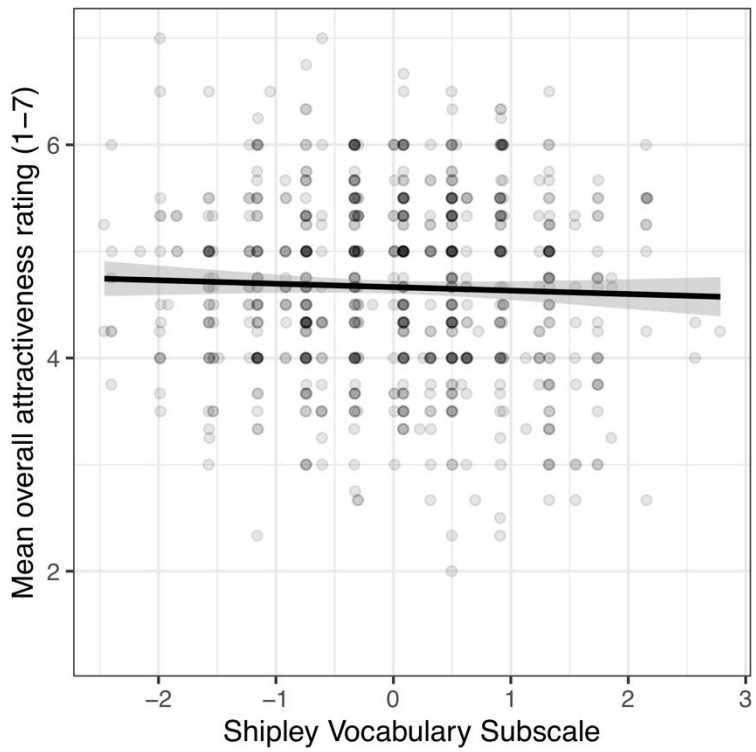


Figure 4: The association between intelligence, as measured by the Shipley Institute of Living Scale (Vocabulary Subscale), and rated mate appeal, after aggregating across raters.

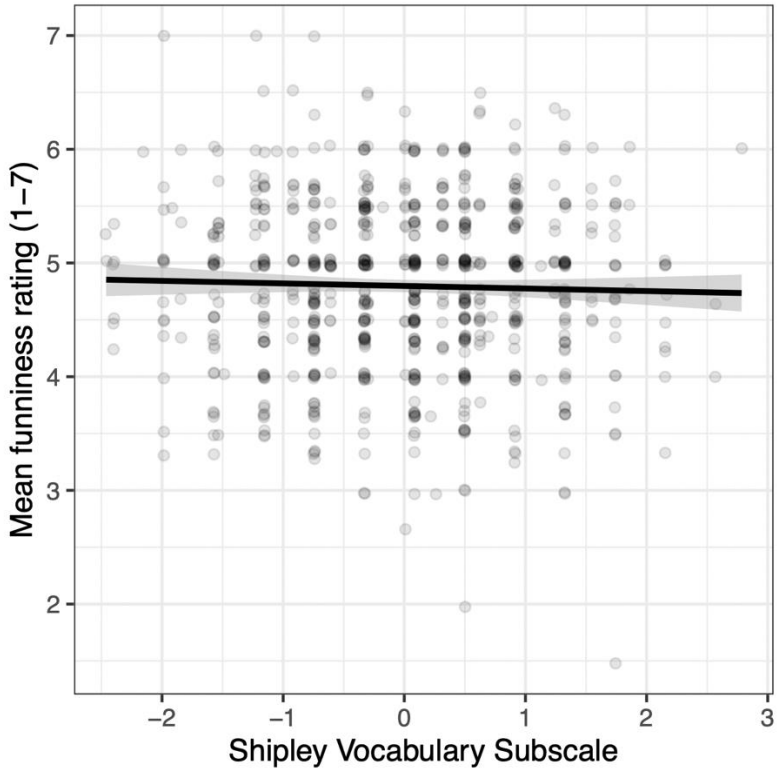


Figure 5: The association between intelligence, as measured by the Shipley Institute of Living Scale (Vocabulary Subscale), and funniness, after aggregating across raters.

Table 3: *MLM coefficients for associations between the Shipley Institute of Living Scale (Vocabulary Subscale), rated intelligence, rated funniness, and rated mate appeal.*

		Rated Mate Appeal (1-7)			
Predictors	*Estimates*	*CI*	*p*	$N_{interactions}$	$N_{participants}$
Shipley (Vocabulary Subscale)	-0.02	-0.07 – 0.04	.560	2114	727
Rated Intelligence	0.29	0.25 – 0.34	<.001	2118	728
Rated Funniness	0.41	0.37 – 0.44	<.001	2118	728

Note: Separate models were used for each predictor. In all models, sex was controlled. Full models are included in supplementary material E.

Discussion

The sexual selection theory of human intelligence proposes that intelligence evolved at least partly as a fitness indicator. Under this scenario, we would expect intelligence to be attractive to members of the opposite sex[2,3]. Past research has established that intelligence is considered a highly attractive trait in a hypothetical long-term mate[8,9]; however, it is less clear whether actual, objectively assessed intelligence is indeed found attractive when evaluating a prospective partner, particularly in a face-to-face context[14,20,29]. Our findings suggest it is not.

First, our results replicate past findings in showing that intelligence is somewhat detectable, particularly in the highly controlled setting of our study 1 (r = .34). This is broadly equivalent to others previously reported in the literature[12,29]. In the more ecologically valid setting of study 2, the association between actual and perceived intelligence is still significant, though attenuated. Taken together, these findings indicate that intelligence can be judged with above chance accuracy by members of the opposite sex at zero acquaintance.

Contrary to our hypotheses, more intelligent people were not rated as having a higher mate appeal by members of the other sex. Instead, only perceived intelligence was associated with higher mate appeal ratings. This finding illustrates the importance of using *measured* intelligence. Given our large sample size, we should have been able to detect an effect between intelligence and mate appeal with the current study. The difference to Prokosch et al. might be best explained by the fact that in their small sample of fifteen target men, the association between mate appeal and intelligence may have been stronger simply due to sampling error. Given that rated physical attractiveness and perceived intelligence were strong predictors of (sexual) mate appeal while measured intelligence was not, a halo effect could play a role. It is well established that physically attractive individuals are perceived as better in other socially desirable domains, whether or not any objective differences exist[30].

By gradually increasing the intelligence information and estimating the effect of intelligence above and beyond what can be observed from only physical cues, we could isolate the effect of information about intelligence, without the halo effect of physical attractiveness or any effects that intelligence might have on cues such as clothing or body shape. Contrary to our hypotheses, the increase in mate appeal after adding intelligence-related cues to visual and vocal attractiveness cues was not enhanced for more intelligent men.

In both studies, we found that ratings of funniness were associated with ratings of mate appeal, but contrary to our hypotheses, measured intelligence did not predict ratings of funniness in either study. According to the sexual selection theory of human intelligence, our ancestors would have used interpersonal humour during courtship to advertise and evaluate underlying intelligence and ultimately genetic quality[2,3]. Our findings contrast with previous work by Greengross and Miller[23] and Howrigan and MacDonald[24] who found that measured intelligence was positively associated with humour production tasks. Our measures of humour in both studies relied on being funny during a live interaction (which was recorded in study 1 and truly live in study 2). Our measures presumably tapped into interpersonal humour, with its real-time evaluation and non-verbal cues, whereas the more abstract tasks by Greengross and Miller[23] and Howrigan and MacDonald[24] may have tapped skills such as drawing and writing that are more related to intelligence but relatively unimportant for interpersonal humour. While our measures had the drawback that participants' ratings of funniness could be contaminated by the halo effect of physical attractiveness, ratings of funniness still predicted mate appeal in both studies after adjusting for rated physical attractiveness.

In both studies, we found that intelligence is detectable to some degree, but that only perceived but not measured intelligence is linked to mate appeal. If perceived intelligence is

found attractive and measured intelligence is not, one possibility is that invalid cues of intelligence are found attractive. Previous research has coded the frequency of different verbal and non-verbal cues and their relationship with measured and perceived intelligence; a cue that is related to perceived intelligence and unrelated to measured intelligence is necessarily an invalid cue. Reynolds and Gifford[17] adopted this technique and showed that a reduction of halting speech and an increase in standard speech were associated with greater perceived intelligence but not with measured intelligence. Similarly, Murphy[13] found that a clear style of speech was positively associated with perceived intelligence but not measured intelligence. As people can detect intelligence to some extent, valid cues of intelligence are clearly perceptible. This is supported by studies examining intelligence cues which found that increased eye contact while speaking is positively associated with both measured intelligence and perceived intelligence[16,17]. Had we evolved to find intelligence attractive because it signals genetic quality, we would have evolved to find valid cues of intelligence attractive. This pattern of results is not consistent with Miller's proposal that intelligence acts as a fitness indicator[2].

Why, then, do people across time and cultures report that intelligence is important in a mate[9,31]? Apart from the possibility that people's conceptions of intelligence are to some extent based on invalid cues, another possible explanation is that intelligence and related constructs are associated with positive outcomes across all environments. Therefore, people in these environments (i.e. cultures) will learn to associate intelligence with positive outcomes and, as a consequence, people will report intelligence as being desirable across cultures. Previous research has shown that people do indeed believe intelligent individuals possess socially desirable traits. Murphy, Hall, and LeBeau[32] showed that people believed intelligent individuals were more competent and open-minded. The association between perceptions of intelligence and socially desirable traits appears to develop early in life, as

children who are asked to draw an intelligent individual, draw high status, socially successful people[33,34]. Intelligence is also linked to more direct benefits related to resource provisioning potential, including socioeconomic status, income, education and health[35], which might imply that intelligence becomes attractive only at later stages of relationship formation[11].

As discussed, we have tried to use complementary approaches in study 1 and study 2, where the limitations of one study are in many cases addressed by the strengths of the other study. Study 1 prioritises precision in our estimates of intelligence and a high degree of control over intelligence information at the expense of ecological validity, whereas study 2 does the opposite. A major limitation of study 1 was that ratings of men's sexual mate appeal were generally low, so that it seems unlikely that many of the men in our sample would have been chosen as partners by our raters and perhaps further reducing ecological validity. In study 2, ratings of mate appeal were higher and many participants indicated hypothetical interest in going on a real date with their partner (43.6% of interactions for women, 47.5% of interactions for men). Another limitation of study 1 was that women only saw short video sequences and did not have any real interaction with the targets. At this initial stage of courtship, physical attractiveness is the most influential; however, it is possible that in later stages intelligence would become more important. This issue is partly addressed in study 2 by having participants interact with each other in an uncontrolled manner that is more reflective of a real courtship situation; however, we are still limited to the initial phase of getting acquainted, which may underestimate the importance of intelligence throughout a courtship. Future studies could address this by adopting a longitudinal design that assesses intelligence and attraction through progressive stages of acquaintance. Still, the fact that our participants could detect intelligence but were not influenced by it in their ratings of mate appeal calls into question the idea that intelligence is a fitness indicator.

A limitation of study 2 was that ratings of intelligence could be contaminated by cues about income, such as clothing and accessories. This limitation is mitigated in study 1 by showing images and voice prior to video content and controlling for these previous ratings when testing for the association between intelligence and mate appeal. Another limitation of study 2 is that our intelligence test only measured vocabulary and is therefore an imprecise estimate of our participants' intelligence. This does not apply to study 1, where a *g* factor based on multiple intelligence tests was calculated, thereby greatly increasing the reliability and validity of the intelligence construct. Relatedly, intelligence scores in study 2 were based on a university sample that is more educated and likely has a higher socioeconomic status than the general population. We partly addressed this in study 1, which was based on individuals from university and the broader population to provide more diverse backgrounds and likely more diverse intelligence scores. However, all targets in study 1 were literate and not intellectually disabled, which means that intelligence variation was still limited to some degree. It is possible intelligence is important in a mate only to the extent that it is not very low[36], which our samples would not detect.

In conclusion, our results do not support Miller's proposal that human surplus intelligence was shaped by intersexual selection. If our intelligence was shaped by the romantic and sexual choices across generations, this legacy should not only be reflected in our stated preferences, but also in who we find attractive as a partner. Instead we found that, at least in initial evaluations of potential partners, measured intelligence did not influence mate appeal, neither directly nor indirectly through funniness. Given the caveats to our findings, future research should extend our work by sampling a broader variation of the spectrum of intelligence and following courtship over a longer term beyond the initial contact.

Method

Study 1

Data collection for study 1 was completed in three steps: an online questionnaire and follow-up lab-based session with male participants (stimuli), and several lab-based sessions with female participants (raters). All participants provided written consent and were informed about the study's aim after participation. Studies like ours are exempt from IRB according to German regulations.

Participants

Male targets. An online survey titled 'Person Perception' was used to screen participants for inclusion in our lab-based study. Participants were recruited with posters in the city centre (e.g. train stations, gyms, job centres) and the Goettingen university campus. Of the 347 participants that commenced the survey, 118 males finished. All of these 188 men over the age of 18 years were then recruited to participate in our lab-based study. Final participants were 88 males with ages ranging from 19 to 31 years (M = 24.22, SD = 2.81). Participation was incentivised through a small payment (10€) and personalised feedback on their personality. The sample varied in educational attainment, ranging from university degrees (26%), high school degrees (67%), vocational baccalaureate diploma (5%), to secondary school leaving certificates (2%). The vast majority of the sample was heterosexual (97%), with one homosexual and two bisexual participants. The majority of men were single (61%) and the remainder were currently in a romantic relationship (39%).

Female raters. Participants were recruited through various online channels (e.g. Facebook, a local student participant pool) and posters on campus. Of the 203 participants that responded, 24 were excluded on the basis of either being male (14), technical difficulties (9), or previous participation (1). We also excluded ratings in which women reported acquaintance with the male target, leaving a final number of 39,003 ratings (3% dropout)

from 179 females with ages ranging from 18 to 36 years (M = 21.84, SD = 3.22). Participation was incentivised through a coupon lottery and course credit for those recruited at the university. The vast majority of the sample was heterosexual (93%), with one homosexual participant (1%), and 11 bisexual (6%) participants; 55% were in relationships and 45% were single.

Participants were distributed across six rating studies (described in greater detail in S2) with the sample size breakdown as follows: study 1.1 (n = 19, ratings = 1657), 1.2 (n = 16, ratings = 1368), 1.3 (n = 30, ratings = 2620), 1.4 (n = 25, ratings = 10,485), 1.5 (n = 30, ratings = 12,739), and 1.6 (n = 59, ratings = 10,134). Demographics for individual groups are reported in the supplementary materials (S2).

Materials and Procedure

Male targets. Participants completed an online questionnaire implemented via the survey framework formr.org[37]. The questionnaire included basic demographic items (e.g. age, gender, sexual orientation, and educational attainment), along with more extensive measures related to intelligence (extended German version of the International Cognitive Ability Resource (ICAR)[38], and personality (irrelevant to the current study). Each subsequent laboratory session, which yielded the stimuli for study 1, lasted approximately one hour and was conducted by the same two female experimenters to standardise experimenter effects across participants and induce potential effects of female presence on male self-display behaviour[39,40].

At the beginning of the session, additional assessments of men's measured intelligence were applied, namely the Deary-Liewald Reaction Time Task (DLRT[41]), the Multiple Choice Vocabulary Test (MWT-B[42]), and the knowledge scale from the Berliner Test zur Erfassung Fluider und Kristalliner Intelligenz (BEFKI GC-K[43]). Men were then photographed and videotaped performing several tasks (see Figure 6). First, a facial

photograph (cue 1) and second a full body photograph (cue 2) of men standing on a marked spot to standardise lighting and focal distance was taken. Men received no instructions for posture and facial expression. Third, we videotaped men reading vowels out loud (cue 3). Each vowel was displayed onscreen for two seconds each to standardise reading speed. Fourth, the men were videotaped while reading five newspaper headlines from German newspapers aloud as this task is strongly related to an accurate intelligence perception[44]. In order for them to be intellectually challenging, we selected headlines containing foreign words or describing complex facts (e.g. 'Compensation payments lead US diocese into bankruptcy.'). Fifth, we videotaped men pantomiming the word 'Zahnrad' (mechanical gear) which we used as a warm-up and the word 'Bankverbindung' (bank connection) (cue 5). Last, men were asked to make the experimenter laugh within a 30 seconds time limit by telling an anecdote or joke (cue 6); they were given five minutes to prepare for this task prior to video recording. Full HD cameras (resulting in a resolution of 1920 x 1080 pixels) were used for all recordings and clips were created with the program Mangold VideoSyncPro IP Version 1.7.0.22[45].

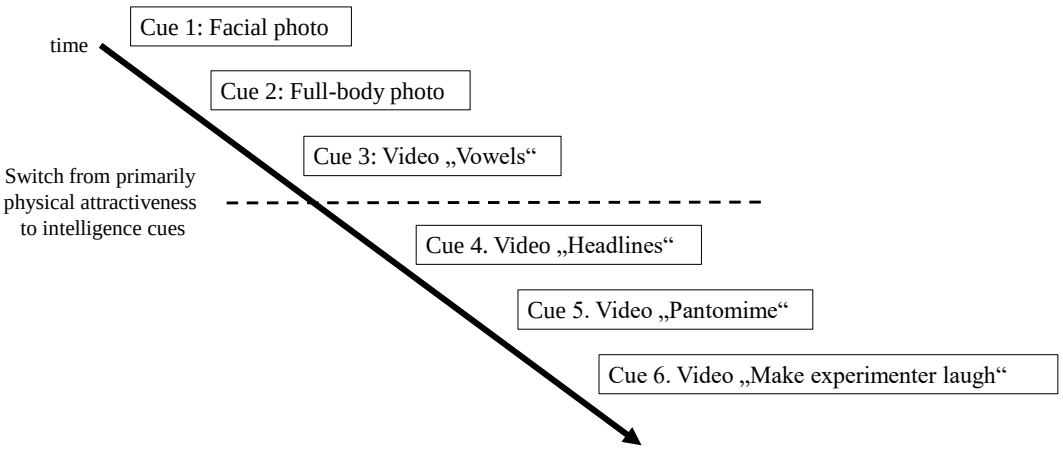


Figure 6. Overview of study 1 stimuli.

Female raters. Female raters participated in one of six computer-based rating studies (referred to herein as rating study 1.1 - 1.6) based on slightly different sets of stimuli. For all rating studies, the session began with a short demographic questionnaire, including age, gender, educational attainment, relationship status, and sexual orientation. Rating study 1.1 assessed a baseline of men's physical attractiveness, 1.2 assessed perceived intelligence and funniness, and 1.3 assessed men's attractiveness as a short-term mate and long-term mate. Rating study 1.4 assessed changes in men's short-term mate attractiveness when shifting from physical attractiveness information (cues 1-3) to additional cues related to men's intelligence (cue 4 and cue 6). Rating studies 1.5 and 1.6 were replications of rating study 1.4 with small methodological improvements. Stimuli were randomised into two blocks: after watching the first block, participants were able to take a 15-minute break to reduce test fatigue. In rating study 1.6, women only rated a randomly drawn half of our target sample (44 men) to further reduce test fatigue; in all other studies, all 88 men were rated. Studies 1.1 to 1.4 were programmed using the Software PsychoPy2 Experiment Builder (v1.80.06)[46]; however, a software update of PsychoPy crashed experiment 1.5, therefore, we ran study 1.6 and the majority of study 1.5 on the experimental framework Alfred[47].

Rating study 1.1. Participants rated the target's physical attractiveness after being shown two photographs (cue 1: facial photograph; cue 2: full body photograph). The item ('How attractive do you find this man?') was rated on a 7-point Likert-scale ranging from 1 (not attractive at all) to 7 (very attractive).

Rating study 1.2. Participants rated targets' intelligence and funniness after watching three video sequences of each target (cue 4-6). The item (e.g. 'He is intelligent', 'He is humorous') was rated on a 5-point Likert-scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Rating study 1.3. Participants watched the same three videos as in rating study 1.2 (i.e. cue 4-6); however, they were instead asked to evaluate men's short term- and long-term mate attractiveness. The items ('How well can you imagine having a sexual affair with this man?' and 'How well can you imagine a long-term relationship with this man?') were rated on a 7-point Likert-scale ranging from 1 (not at all) to 7 (very well).

Rating study 1.4. Participants were provided with definitions of short-term mate ('brief sexual encounters') and long-term mate ('serious, committed relationships') prior to ratings. Participants provided ratings five times: first after they saw facial photographs (cue 1), then after seeing full body photographs (cue 2), then after seeing each of three additional videos (cue 3, 4, and 6). Cue 5 was not presented in order to reduce test fatigue. Each time the item ('Please rate the following recording of this man considering his short-term and long-term mate attractiveness') was rated on two separate response scales ranging from 1 (not attractive) to 100 (very attractive).

Rating study 1.5. The procedure for rating study 1.5 was almost identical to rating study 1.4; however, participants were now instructed to evaluate men's short-term and long-term mate attractiveness independently of their own relationship status. That is, women were asked to provide ratings from the perspective of a single woman even if they were partnered. Additionally, women saw a preview of all 88 facial photographs of the target men prior to making any responses. These modifications were made because the ratings in the first study were extremely low (mean of 19 on a scale from 0 to 100), suggesting a floor effect. By previewing the full range of men in the study, we hoped that women would not reserve their highest attractiveness rating in the expectation that a more attractive man would appear. For the preview, each man's picture was displayed for two seconds in a randomised order. As a final attempt to improve discrimination between targets, we also explicitly pointed out the whole range of the scale to participants.

Rating study 1.6. The procedure for rating study 1.6 slightly improved upon rating study 1.5 with an aim of reducing potential fatigue effects. In this study, twice the number of female raters rated half of the targets (44 of 88). Additionally, women saw men's facial and full body photographs (cue 1 and 2, respectively) and made their first rating based on both photos. The items were phrased identically to rating studies 1.4 and 1.5; however, the scale now ranged from -50 (repulsive) to +50 (attractive). The slider was preset to the scale's midpoint (0).

Statistical Analyses

All our analyses were run using R 3.6.0[48].

Male stimuli. Targets' measured intelligence, extracted as a *g* factor, is the first unrotated factor of a principal component analysis of the eight intelligence tests used in study 1.

Accuracy of intelligence perception. For each male target, we aggregated all women's ratings of men's intelligence to calculate the *aggregated perceiver accuracy*. We correlated men's actual intelligence with this aggregated perceived intelligence to investigate the accuracy of intelligence perception. Additionally, we fitted a structure equation model in *lavaan* v0.6-4[49] modelling *g* as a hierarchical latent variable to correct for measurement error and clustering standard errors by target to estimate the semi-latent single rater accuracy.

Preference for Intelligence. To test whether intelligence adds a unique contribution to men's long-term and short-term mate attractiveness, we used Bayesian multilevel linear models calculated in Stan[50] with the *brms* package v 2.10.0[51] with weakly informative priors. To validate our analyses, we additionally fitted models in *lme4* v1.1-21[52]. As ratings resulted from three different studies (rating studies 1.4-1.6), we included an interaction between study and each cue, allowing for varying influences of cues on long-term mate/short-term mate ratings in each study. Because the studies grouped cues differently, the cue variable had four

levels that were entered dummy-coded: face/body photo, voice, newspaper headlines and make experimenter laugh, with the voice recording set as the reference category. Of main interest, we specified population-level interactions between the cues and intelligence. These were adjusted for by specifying interactions between cues and physical attractiveness. We specified varying intercepts for targets and raters. Additionally, we allowed the effect of the cue dummy variables to differ between targets and the interaction between cues and traits to differ by rater. Finally, we let an interaction between cue and study and varying intercepts for raters and targets predict the residual standard deviation in the regression in a location-scale model to account for the fact that the rating scale might be used differently across studies and participants.

Preference for funniness. To assess the influence of funniness incremental to the influence of measured intelligence on mate appeal, we regressed men's *g* factor and ratings of their funniness onto their mate appeal. We used the packages *sandwich* v2.5-1[53,54] and *lmtest* v0.9-37[55] to correct our standard errors as ratings of men's physical attractiveness, funniness and mate appeal were clustered in three different sets of female raters.

Robustness Checks. We stated in the preregistration that we would only recruit heterosexual raters and targets, so we repeated all of our analyses excluding participants who indicated that they were not heterosexual. We also stated in the preregistration that we would use aggregated ratings instead of women's individual ratings for a given trait. Those aggregations were planned for physical attractiveness, short-term mate attractiveness, long-term mate attractiveness, perceived intelligence and perceived funniness. We repeated these analyses as a robustness check.

Study 2

Participants

Participants were 729 (379 female) first year psychology students with ages ranging from 16.92 to 36.08 years (females: M = 19.24, SD = 2.64; males: M = 19.74, SD = 2.64). Participants were recruited between 2016 and 2019 from the University of Queensland's first year research participation scheme and were offered one credit for their participation in a study titled 'Speed-meeting Study'. To participate in the study, participants were requested to be 1) heterosexual, 2) a native English speaker, 3) open to answering personal questions regarding their sexual history (for questions not relating to the current study), and 4) not in a committed relationship (required in 2017-2019). Participants who were known to each other (3.80%) or in a committed relationship (7.30%) were included in the main analyses; however, results with these participants excluded can be found in supplementary material E. Participants said yes to going on another date with their partner 46.54% of the time and they mutually said yes 20.95% of the time.

Before beginning, all participants were asked to read an information sheet which briefly detailed the procedure and highlighted the potential sensitivity of the sexually oriented questions. Participants were assured of confidentiality as well as being told at regular intervals that they may discontinue/omit answers without forgoing credit. They were then given an educational debriefing, including a debrief sheet. This study was approved by the Human Ethics Committee at the University of Queensland (Ethics #16-PSYCH-4-65-JS).

Materials

Participants completed a series of questionnaires that were collected as part of a larger study investigating attraction. Only items included in the present study are detailed below.

Demographics. A range of demographic questions including age, sex, sexual orientation, and relationship status.

Speed-date ratings. Participants completed a 24-item questionnaire regarding each partner with whom they had a speed-date interaction. The first series of questions concerned the partner's personality attributes. Participants were asked to 'Please rate this partner on the following statements below' and were then presented with a statement regarding each trait individually, such as, for example, 'They are funny'. To ensure participants paid attention to the intelligence trait in particular, it was separated from the other traits and asked in the longer format of 'Thinking about this interaction, approximately how intelligent do you think this partner is?' The second series of questions concerned the partner's facial, bodily, and overall attractiveness (e.g. 'I would rate their overall attractiveness as…'). All questions in this section were rated on a 7-point response scale ranging from 1 = *Well Below Average* to 7 = *Well Above Average* with a midpoint of 4 = *Average*.

Verbal intelligence. To measure verbal intelligence, the latter (more difficult) half of Shipley's Vocabulary Scale was used[56]. This scale included 20 items whereby the participant is presented with a target word (e.g. 'Jocose') and a series of four words (e.g. 'Humorous, Paltry, Fervid, Plain'). Participants are instructed that for each target word, they should 'please select the word that best matches its meaning'. These items progressively become more difficult, beginning with well-known words such as 'Caption' and ending with more obscure words such as 'Temerity'.

Procedure

Pre-date. Four speed-date stations were constructed in the laboratory. Participants were seated opposite each other with Apple iPads so they were unable to see their partner's screen. Each station was separated by 1.7m room dividers to ensure the other couples were also unable to see their device screens. Upon arrival, participants were seated and given a participant information sheet. They were instructed to begin the pre-questionnaire if they agreed to participate. The pre-questionnaire consisted of demographics and other measures

not used in this study. At the end of the pre-questionnaire, participants received on-screen instructions to wait quietly until others were finished.

Speed-dating. Once all participants had completed the pre-questionnaire, they were verbally instructed that they would now be given three minutes to interact with an opposite sex partner. Participants were instructed to speak about any topic until they heard a bell which would indicate the date had ended. After hearing the bell, participants were then instructed to begin completing the survey regarding their partner (as outlined in the speed-date ratings section of Materials). All participants were reminded to hold the iPads up to avoid their partner seeing the screen. Experimenters supervised the room to determine when all participants had finished completing ratings. At that point, the rotating sex (counterbalanced) moved onto the next station to start their next date. The process was then repeated until all opposite-sex dyads had interacted. If there was an uneven ratio of males and females, the extra participant(s) were instructed to sit quietly for three minutes during that round. In total, there were 123 speed-dating sessions with 729 participants. Participants participated in 2-5 dates (M = 3.01).

Post-date. Once all speed-dates and ratings had been completed, participants began completing the post-questionnaire which consisted of Shipley's Vocabulary Scale[56]. Participants completed the first two sections and were instructed to wait quietly until all others had finished.

Statistical analysis

The nature of the design (i.e. participants rating multiple partners) creates dependencies in the data. The rating from each interaction between two people (Level 1) is cross-classified within both the participant receiving the rating (Level 2), and the partner who gave the rating (Level 2), all of which is nested within the session they both attended (Level 3). Therefore, it is necessary to use multilevel modeling (MLM) to account for the

hierarchical structure of the data. MLM analyses with partner ratings of attractiveness and intelligence at Level-1 and measured intelligence at Level-2 were used to evaluate main effects. Additionally, random slopes were included for all main effect variables (e.g. measured intelligence) for both grouping factors (i.e. participant, partner) to allow the slope between the independent and dependent variable to vary by group; however, these random slopes were removed when necessary to resolve convergence issues.

References

1. Mink, J. W., Blumenschine, R. J. & Adams, D. B. Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis. *The American Journal of Physiology* **241,** R203 - R212; 10.1152/ajpregu.1981.241.3.R203 (1981).
2. Miller, G. *The mating mind. How sexual choice shaped the evolution of human nature* (Doubleday, New York, NY, 2000).
3. Miller, G. F. in *Evolutionary perspectives on human reproductive behavior*, edited by D. LeCroy & P. Moller (New York Academy of Sciences, New York, 2000), pp. 62–74.
4. Darwin, C. *The descent of man and selection in relation to sex* (John Murray, London, 1871).
5. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489,** 391–399; 10.1038/nature11405 (2012).
6. Klasios, J. Cognitive traits as sexually selected fitness indicators. *Review of general psychology* **17,** 428–442; 10.1037/a0034391 (2013).
7. Puts, D. A. Beauty and the beast. Mechanisms of sexual selection in humans. *Evolution and Human Behavior* **31,** 157–175; 10.1016/j.evolhumbehav.2010.02.005 (2010).
8. Li, N. P., Bailey, J. M., Kenrick, D. T. & Linsenmeier, J. A. W. The necessities and luxuries of mate preferences. Testing the tradeoffs. *Journal of Personality and Social Psychology* **82,** 947–955; 10.1037/0022-3514.82.6.947 (2002).
9. Buss, D. M. *et al.* International Preferences in Selecting Mates. A Study of 37 Cultures. *Journal of Cross-Cultural Psychology* **21,** 5–47 (1990).
10. Eastwick, P. W., Eagly, A. H., Finkel, E. J. & Johnson, S. E. Implicit and explicit preferences for physical attractiveness in a romantic partner. A double dissociation in predictive validity. *Journal of Personality and Social Psychology* **101,** 993–1011; 10.1037/a0024061 (2011).
11. Miller, G. F., & Todd, P. M., Mate choice turns cognitive. *Trends in cognitive sciences*, **2**, 190–198; 10.1016/S1364-6613(98)01169-3 (1998).
12. Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M. & Angleitner, A. Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology* **86,** 599–614; 10.1037/0022-3514.86.4.599 (2004).
13. Murphy, N. A. Appearing smart. The impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personality & social psychology bulletin* **33,** 325–339; 10.1177/0146167206294871 (2007).
14. Scheib, J. E., Gangestad, S. W. & Thornhill, R. Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266,** 1913–1917 (1999).
15. Borkenau, P. & Liebler, A. Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology* **65,** 546–553; 10.1037/0022-3514.65.3.546 (1993).
16. Murphy, N. A., Hall, J. A. & Colvin, C. R. Accurate intelligence assessments in social interactions. Mediators and gender effects. *Journal of Personality* **71,** 465–493; 10.1111/1467-6494.7103008 (2003).

17. Reynolds, D.'A. J. & Gifford, R. The Sounds and Sights of Intelligence. A Lens Model Channel Analysis. *Personality & social psychology bulletin* **27,** 187–200; 10.1177/0146167201272005 (2016).

18. Carney, D. R., Colvin, C. R. & Hall, J. A. A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality* **41,** 1054–1072; 10.1016/j.jrp.2007.01.004 (2007).

19. Prokosch, M. D., Coss, R. G., Scheib, J. E. & Blozis, S. A. Intelligence and mate choice. Intelligent men are always appealing. *Evolution and Human Behavior* **30,** 11–20; 10.1016/j.evolhumbehav.2008.07.004 (2009).

20. Sprecher, S. & Regan, P. C. Liking Some Things (in Some People) more than Others. Partner Preferences in Romantic Relationships and Friendships. *Journal of Social and Personal Relationships* **19,** 463–481; 10.1177/0265407502019004048 (2002).

21. Bressler, E. R. & Balshine, S. The influence of humor on desirability. *Evolution and Human Behavior* **27,** 29–39; 10.1016/j.evolhumbehav.2005.06.002 (2006).

22. Wilbur, C. J. & Campbell, L. Humor in romantic contexts. Do men participate and women evaluate? *Personality & social psychology bulletin* **37,** 918–929; 10.1177/0146167211405343 (2011).

23. Greengross, G. & Miller, G. Humor ability reveals intelligence, predicts mating success, and is higher in males. *Intelligence* **39,** 188–192; 10.1016/j.intell.2011.03.006 (2011).

24. Howrigan, D. P. & MacDonald, K. B. Humor as a Mental Fitness Indicator. *Evol Psychol* **6,** 147470490800600; 10.1177/147470490800600411 (2008).

25. Ambady, N. & Rosenthal, R. Thin slices of expressive behavior as predictors of interpersonal consequences. A meta-analysis. *Psychological Bulletin* **111,** 256–274; 10.1037/0033-2909.111.2.256 (1992).

26. Plomin, R. & Deary, I. J. Genetics and intelligence differences. Five special findings. *Molecular psychiatry* **20,** 98–108; 10.1038/mp.2014.105 (2015).

27. Back, M. D. & Nestler, S. in *The Social Psychology of Perceiving Others Accurately*, edited by Hall, Judith A. and Schmid Mast, Marianne and West, Tessa V. (Cambridge University Press, Cambridge, 2016), pp. 98–124.

28. Milonoff, M. & Nummi, P. Adolescents but not Older Women Misjudge Intelligence from Faces and Do not Consider Intelligent-Looking Men Attractive. *Annales Zoologici Fennici* **49,** 378–384; 10.5735/086.049.0512 (2012).

29. Gangestad, S. W., Garver-Apgar, C. E., Simpson, J. A. & Cousins, A. J. Changes in women's mate preferences across the ovulatory cycle. *Journal of Personality and Social Psychology* **92,** 151–163; 10.1037/0022-3514.92.1.151 (2007).

30. Langlois, J. H. *et al.* Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin* **126,** 390–423; 10.1037/0033-2909.126.3.390 (2000).

31. Souza, A. L., Conroy-Beam, D. & Buss, D. M. Mate preferences in Brazil. Evolved desires and cultural evolution over three decades. *Personality and Individual Differences* **95,** 45–49; 10.1016/j.paid.2016.01.053 (2016).

32. Murphy, N. A., Hall, J. A. & LeBeau, L. S. Who's smart? Beliefs about the expression of intelligence in social behavior. *Representative Research in Social Psychology* **25,** 34–42 (2001).

33. Räty, H. *et al.* Portraying intelligence. Children's drawings of intelligent men and women in Finnish and Russian Karelia. *Educational Studies* **38,** 573–586; 10.1080/03055698.2012.661928 (2012).

34. Raty, H. & Snellman, L. Children's images of an intelligent person. *Journal of Social Behavior and Personality* **12,** 773–784 (1997).

35. Deary, I. J., Intelligence. *Annual Review of Psychology*, **63**, 453–482; 10.1146/annurev-psych-120710-100353 (2012).

36. Zebrowitz, L. A., & Rhodes, G. Sensitivity to "bad genes" and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of nonverbal behavior*, **28**, 167-185; 10.1023/B:JONB.0000039648.30935.1b (2004).

37. Arslan, R. C., Tata, C. S. & Walther, M. P. *formr: A study framework allowing for automated feedback generation and complex longitudinal experience sampling studies using R* (Behavior Research Methods, 2019).

38. Condon, D. M. & Revelle, W. The International Cognitive Ability Resource. Development and initial validation of a public-domain measure. *Intelligence* **43,** 52–64 (2014).

39. Ronay, R. & Hippel, W. v. The Presence of an Attractive Woman Elevates Testosterone and Physical Risk Taking in Young Men. *Social Psychological and Personality Science* **1,** 57–64; 10.1177/1948550609352807 (2010).

40. Roney, J. R. Effects of visual exposure to the opposite sex. Cognitive aspects of mate attraction in human males. *Personality and Social Psychology Bulletin* **29,** 393–404 (2003).

41. Deary, I. J., Liewald, D. & Nissan, J. A free, easy-to-use, computer-based simple and four-choice reaction time programme. The Deary-Liewald reaction time task. *Behavior Research Methods* **43,** 258–268 (2011).

42. Lehrl, S. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B.* 5th ed. (Spitta Verlag, Balingen, 2005).

43. Schipolowski, S. *et al.* BEFKI GC-K: eine Kurzskale zur Messung kristallinger Intelligenz. *Methoden, Daten, Analysen (mda)* **7,** 153–181; 10.12758/mda.2013.010 (2013).

44. Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M. & Angleitner, A. Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology* **86,** 599–614; 10.1037/0022-3514.86.4.599 (2004).

45. Mangold, P. *Mangold INTERACT Software* (Mangold International GmbH, Arnstorf, Germany).

46. Peirce, J. W. PsychoPy--Psychophysics software in Python. *Journal of neuroscience methods* **162,** 8–13; 10.1016/j.jneumeth.2006.11.017 (2007).

47. Treffenstaedt, C. & Wiemann, P. *Alfred - a library for rapid experiment development* (2018).

48. R Core Team. *R: A language and environment for statistical* (R Foundation for Statistical Computing, Vienna, Austria, 2019).

49. Rosseel, Y. lavaan. An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48,** 1–36 (2012).
50. Carpenter, B. *et al.* Stan. A probabilistic programming language. *Journal of Statistical Software* **76,** 1–32; 10.18637/jss.v076.i01 (2017).
51. Bürkner, P. brms. An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **80,** 1–28; 10.18637/jss.v080.i01 (2017).
52. Bates D., Mächler M., Bolker B. & Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67,** 1–48; 10.18637/jss.v067.i01 (2015).
53. Zeileis, A. Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software* **11,** 1–7; 10.18637/jss.v011.i10 (2004).
54. Berger, S., Graham, N. & Zeileis, A. *Various Versatile Variances. An Object-Oriented Implementation of Clustered Covariances in R* (2017).
55. Zeileis, A. & Hothorn, T. Diagnostic Checking in Regression Relationships. *R News* **2,** 7–10 (2002).
56. Zachary, R. A., & Shipley, W. C. (1986). *Shipley institute of living scale: Revised manual*: WPS, Western Psychological Services.

Acknowledgements

We thank Julia Specht, Christina Hillebrand, Philipp Grochowski and Pia Drewke for help collecting the data of study 1. We thank Peter Zezula for his support setting up parts of study 1.

Author Contributions

JMH, LP, RCA planned and supervised study 1. JCD and MD conducted parts of study 1. RCA and JCD analysed study 1 with consultation from PCB. BPZ and MJS planned study 2. MJS partially conducted, supervised student data collection for, and analysed study 2. MJS and JCD wrote the manuscript. All authors read and approved the final manuscript.

Competing Interests statement

All authors declare no competing interests.

Supplement

S1. Table with deviations from pre-registration

Study 1 was pre-registered as part of three different theses. All theses had a slightly different focus. After the preregistration, the project continued to develop and we collected additional rating data to address methodological shortcomings. Throughout the process, we realised that some preregistered hypotheses were insufficient. We also opted for more appropriate analyses. In the following all deviations from the pre-registration are outlined.

Deviations in our hypotheses

Preregistration	**Manuscript**	**Explanation/Solution**
Thesis A focuses on the association between intelligence and attractiveness and made the following predictions: A1. More intelligent men are preferred as long-term mates by women. A2. Men who are perceived as more intelligent are preferred as long-term mates by women. **A3. More intelligent men are preferred as short-term mates by women.** **A4. Men who are perceived as more intelligent are preferred as short-term mates by women.** A5. Perceived creativity is predicted by intelligence and perceived intelligence. A6. Men who are perceived as more creative are preferred as long-term mates, incremental to intelligence.	We derived the following hypotheses in our manuscript: 1. Women's intelligence ratings for male targets will be positively correlated with targets' psychometrically measured intelligence. 2. Men's measured intelligence will be significantly positively correlated with women's attractiveness ratings. 3. Perceived funniness is associated with measured intelligence. Men's funniness and men's perceived intelligence predicts their rated mate appeal above and beyond the effect of their intelligence.	All preregistered hypotheses in bold are still part of our manuscript (although phrased differently). Differences are: 1. All hypothesis on long-term mate ratings are not part of our main manuscript anymore, however reported in our supplement. This deviation results from the fact that short-term mate and long-term mate ratings were highly correlated (r = .92). In our design participants do not seem to differentiate much between short-term mate and long-term mate ratings. Hence, we cannot test for differences between the two outcomes. 2. None of the hypotheses of thesis C are included in the current manuscript. Thesis C focused on the accuracy of personality

Preregistration	Manuscript	Explanation/Solution
A7. Men who are perceived as more creative are preferred as short-term mates, incremental to intelligence. **A8. Perceived sense of humor is predicted by intelligence and perceived intelligence.** A9. Men who are perceived as more humorous are preferred as long-term mates, incremental to intelligence. **A10. Men who are perceived as more humorous are preferred as short-term mates, incremental to intelligence.** **A11. Women can accurately assess men's intelligence based on thin slices of behavior.** A12. Narcissism and shyness may moderate the effect of intelligence on men's appeal to women as long- and short-term mates, as they are expected to have an effect on intelligence displays and their perception at zero acquaintance independent of actual target intelligence. Similarly, thesis B focuses on the association of intelligence and attractiveness. Though, the thesis goes beyond the former in a more detailed investigation of the relationship between the two. B1. There will be a significant change in the rating of men´s attractiveness as short-term and long-term mates after shifting from mere visual and vocal attractiveness information (full body photo, video	4. The increase in men's short-term mate attractiveness after adding additional cues related to intelligence (i.e. reading newspaper headlines; making experimenter laugh) will depend on men's intelligence, such that the attractiveness increase will be greater for more intelligent men.	judgments and do not fit the scope of this paper. 3. Similarly, we did not include any hypothesis on Narcissism, shyness and creativity for the sake of brevity and since we found no main effect of intelligence on attraction.

Preregistration	Manuscript	Explanation/Solution
of reading vowels aloud) to cues also indicative of intelligence (reading headlines aloud, telling experimenter something funny). Effects are expected to be more pronounced when it comes to long-term mating, but if they are also found for short-term mating this can be interpreted as evidence for intelligence as a cue to genetic quality. **B2. The more cues indicative of intelligence are added, the larger the change in appeal ratings is expected to be. Appeal ratings should increase more for more intelligent men, and more so after the fourth (telling something funny) than after the third rating (headlines).** These intelligence-dependent increases should occur for both long-term and short-term mate ratings, but more so when it comes to long-term mating. **B3. More intelligent men will be rated as more desirable short-term and long-term mates.** Thesis C focused on the accuracy of intelligence and personality judgements. C1. There will be a positive correlation between self-reports and observer ratings of the Big Five dimensions. Correlations are expected to range between r = 0.2 to 0.4. C2. The correlations between self-reports and observer ratings will be higher for extraversion		

Preregistration	Manuscript	Explanation/Solution
and conscientiousness than for neuroticism, openness to experience, and agreeableness C3. Measured psychometric intelligence and self-rated openness to experience are strong predictors for observer-rated creativity and observer-rated humour production ability. C4. Attempting to replicate the results of a study by Murphy and colleagues (2003), the ratings of intelligence made by female raters are predicted to be more accurate (more highly correlated with measured psychometric intelligence) than the ones made by male raters.		

Deviations in our recruitment

Preregistration	Manuscript	Explanation/Solution
Recruitment of targets was limited: • to 80 male participants • men who report a heterosexual orientation • who are aged between 18 - 30	Recruitment slightly differed: • 88 target men were recruited • 2 target men reported a bisexual orientation and 1 target reported to be homosexual • age ranged from 18 to 31 years	In a robustness check, we excluded targets reporting a non-heterosexual orientation targets. Results are reported in our appendix S3.
Recruitment of 55 female raters: • who report a heterosexual orientation • 10 women rate men's physical attractiveness (Rating Study 1.1)	90 [179] female rater were recruited: • 1 reported a bisexual orientation • 19 women participated in Rating Study 1.1 (physical attractiveness)	A total of 179 female raters were recruited. At the time point when writing the preregistration, only Rating Study 1.1 - 1.4 were intended to take place. We decided to run Rating Study 1.5 and 1.6 post-hoc to address methodological shortcomings of study 1.4. Hence, only Rating Study 1.1 - 1.4 should be taken into account

Preregistration	Manuscript	Explanation/Solution
as well as men's humour and intelligence (Rating Study 1.2) • 10 women rate men's STM and LTM attractiveness (Rating Study 1.3) • 25 women rate men's STM and LTM after each presented cue (Rating Study 1.4)	• a different set of 16 women participated in Rating Study 1.2 (rating of intelligence and humour) • 30 women participated in Rating Study 1.3 (LTM and STM attractiveness) • 25 women participated in Rating Study 1.4 (STM/LTM after each Cue) • 30 women participated in Rating Study 1.5 (40 initially participated but 10 had to be excluded due to technical issues) • 59 women participated in Rating Study 1.6	when comparing differences between the manuscript and the preregistration. Thus, the number of preregistered to actually recruited female raters differs from 55 to 90. • It was preregistered that 10 women rate men's physical attractiveness, as well as men's intelligence and humour (Rating Study 1.1 and 1.2). To prevent potential halo effects, we recruited two different sets of female raters. One set rated men's physical attractiveness and the other set rated men's humour and intelligence (described in further detail under deviation in design) • The one woman reporting a bisexual orientation only rated men's intelligence and humour and made no attractiveness rating. Hence, we decided that we do not have to exclude her in any analysis.

Deviations in our design

Preregistration	Manuscript	Explanation/Solution
Among other measures, 24 items of the ICAR (International Cognitive Ability Resource; Condon & Revelle, 2014; German translation by our lab) will be used to measure the targets level of intelligence.	Instead of 24, a total of 25 items out of the ICAR (International Cognitive Ability Resource; Condon & Revelle, 2014; German translation by our lab) were adopted.	We included an additional item of the dimension verbal reasoning in order to increase reliability.
A full body and a facial photograph of each participant will be taken as a stimulus including cues on their physical	Participants did not receive any instruction for posture or facial expression.	Targets did not receive an instruction to have a neutral facial expression and posture due to a miscommunication

Preregistration	Manuscript	Explanation/Solution
appearance (face, body, posture, and style). These will be standardized for posture and neutral facial expression, but in normal street appearance.		with our research assistants. However, we selected the most neutral picture out of our videos which led to semi-standardised pictures.
Female raters watch all three videos of a man and rate his physical attractiveness, intelligence and humour.	To rate men's physical attractiveness, female raters only saw a men's full body and facial picture. A different set of female raters rated men's intelligence and humour based on all three videos of them performing several tasks (video headlines, pantomime and make experimenter laugh).	To minimize potential halo-effects, one set of women rated men's physical attractiveness and a different set of women rated men's intelligence and humour. Physical attractiveness was rated based on the photographs and not on the videos since this is the standard procedure in the literature.

Deviations in our analyses

Preregistration	Manuscript	Explanation/Solution
To test whether women can accurately perceive men's intelligence (hypothesis 1), it was preregistered to correlate men's *g* factor with an aggregated value of perceived intelligence. Female ratings of each men are aggregated.	We correlated the aggregated value of perceived intelligence with men's actual intelligence. Additionally, we calculated the single perceiver accuracy. In a multi-level model, we predicted perceived intelligence ratings with men's actual intelligence. We specified a random effect for each men and each female rater.	Aggregated values inflate the accuracy because measurement error is reduced and because of a wisdom of crowds effect. Because people may mainly have only their own perception to go on, we also calculated the single perceiver accuracy (disattenuated for measurement error in the *g* factor, but not for measurement error in the rating).

Preregistration	**Manuscript**	**Explanation/Solution**
To test whether more intelligent men are rated as more attractive as a STM mate (hypothesis 2), it was preregistered to regress men's actual intelligence onto their STM attractiveness while adding target's age, relationship status as well as experiment (dummy coded) as covariates onto the model. In a second model, instead of actual intelligence the aggregated perceived intelligence was added to the model.	Our main model (explained below under hypothesis 4) allows to answer the question whether more intelligent men are rated as more attractive as a potential partner. A positive main effect of *g* factor would be in line with our second hypothesis. In our supplement we included further analyses on hypothesis 2 based on ratings of rating study 1.3. In these analyses, we specified a linear model and predicted men's STM attractiveness with men's *g* factor, physical attractiveness. The difference is that we now used single ratings. However, we needed to correct standard errors as ratings were clustered in different sets of female raters. In a further model we repeated our analyses just like preregistered using aggregated ratings. We did not include target's relationship status, age and the experimenter as covariates into our model. Nevertheless, omitting these covariates did not change the results.	These analyses were only specified for Rating Study 1.3. Though, we also have the possibility to investigate the assumption in our analyses of Rating Study 1.4 - 1.6 without the need of specifying a further model. To keep our main analyses lean, we only reported the main effect of men's *g* factor in our main model (based on sample 1.4 - 1.6). We shortly mention results of sample 1.3 in the main text but included a more detailed overview our appendix. Nevertheless, the preregistered model relies on aggregated ratings which inflate type I errors (DeBruine, 2019; Judd et al., 2016).
To test the influence of humour on attractiveness as a potential partner above the influence of intelligence (hypothesis 3), again it was preregistered to use aggregated ratings.	We included the preregistered model in our appendix. However, we specified a multilevel model in our manuscript.	We argue a multilevel model instead of aggregating ratings being the more appropriate way of analysis (as explained above).

Preregistration	Manuscript	Explanation/Solution
To test whether the STM attractiveness increases for more intelligent men after presenting additional cues related to intelligence (hypothesis 4), a within-subject ANCOVA was preregistered. While measured intelligence should be added as a covariate to the model, target age, relationship status as well as experimenter should be added as control variables.	We specified Bayesian models with weakly informative priors (for a detailed description see Method and S2C).	At the time of the preregistration, we did not intend to replicate our results in two further studies. However, our analyses have to take into account the varying influences of each rating study, explaining additional differences in our preregistered and actual analyses. Additionally, we no longer consider an ANCOVA to be an appropriate way of analysing our data. A Bayesian model better satisfies the needs of our hierarchical data.

S2. Detailed description of Study 1

A. Materials

Intelligence measures. We adopted multiple measures to assess men's intelligence.

In the online screening survey, we used the 16-item short version of the International Cognitive Ability Resource (ICAR, Condon & Revelle, 2014) which is a public-domain assessment tool to assess cognitive abilities. We enriched the short-version with 9 additional items of the long version to increase reliability. Hence, we assessed 4 dimensions namely verbal reasoning, matrix reasoning, letter and number series as well as mental rotation three-dimensional.

In the lab, we used three additional measures. We adopted the multiple choice vocabulary test [Mehrfachwahl-Wortschatz-Intelligenztest] (MWT, Lehrl, 2005) which is a measure to assess participants' general intelligence level, especially their crystallized intelligence.

Our third measure BEFKI GC-K [Kurzskala des Berliner Tests zur Erfassung fluider und kristalliner Intelligenz] (BEFKI GC-K, Schipolowski et al., 2013) also assesses participants' crystallized intelligence based on a 12-item knowledge scale.

Additionally, we adopt the Deary-Liewald reaction time task which is a computer-based reaction time programme (DLRT, Deary et al., 2011). The DLRT assesses simple reaction times (SRT) as well as four-choice reaction times (CRT). To assess the SRT, in each of the 20 runs participants pressed a button in response to a single stimulus. For the CRT, 4 stimuli were presented in 40 runs. In each run, participants had to press one button corresponding to the correct stimuli.

Table S2.
Mean values, standard deviations and ranges of intelligence measures

Variable	*M*	*SD*	min	max
BEFKI	9.52	1.81	4	12
CRT	450.90	47.15	351.2	557.1
SRT	289.8	21.38	244.60	337.20
ICAR letter	4.38	1.80	0	6
ICAR matrix	3.52	1.50	0	6
ICAR rotation	3.53	2.11	0	6
ICAR verbal	5.21	1.43	1	7
MWT	23.57	4.25	11	32

Note. Sum scores are reported for the MWT.

B. Demographics

Study 1 comprises 6 rating studies which are described in more detail below. Participation was rewarded with course credit. Women could only participate in one of the six rating studies.

Rating study 1.1. In October and November 2014, we assessed the target's physical attractiveness. Hence, 19 women rated men's psychical attractiveness after seeing a facial and body photograph of our male targets (stimuli 1 and 2). Women were on average 23 years old (SD = 3.14, range = 18 - 28). 18 women reported to be students. All women reported a heterosexual orientation. 42.11% women were in a relationship.

Rating study 1.2. In October 2014, we also invited 30 participants who rated the targets' intelligence, funniness, creativity and personality after watching 3 videos of target men (cue 4 - 6). 16 of those raters were female (mean age = 21.06, SD = 3.44, range = 19 - 30) and 14 raters were male (mean age = 21.86, SD = 2.83, range = 19 - 29). For our purposes, only female ratings of men's intelligence and funniness are used in subsequent analysis. 15 women's highest level of education was a high school degree and 1 woman reported a university degree as her highest level of education. 87.5% of those women were in a relationship. 15 women reported a heterosexual orientation and 1 woman reported to be bisexual.

Rating study 1.3. In November 2014, we assessed participants' short-term and long term mate attractiveness. A new set of 30 women rated men's attractiveness as a short-term and long-term mate after watching 3 videos of target men (cue 4 - 6). All women were students and were on average 20.87 years old (SD = 2.42, range = 18 - 28). 36.67% of those women were in a relationship. All women reported to be heterosexual.

Rating study 1.4. From August to September 2014, 25 heterosexual women participated in this study (mean age = 23.96 years, SD = 2.82, range = 20 - 30). 60% reported having a high-school degree and 40% a university degree as their highest level of education. 60% of those women were currently involved in a romantic relationship.

Rating study 1.5. The first replication of study 1.4 took place between June 2015 and August 2015. We recruited through various online channels (e.g. Facebook, university platform), as well as posters on campus at the University of Goettingen. We aimed to recruit 30 raters. Participants had to be female and between 18 and 30 years old. Women had the possibility to either receive course credit or participate in a lottery as an incentive for taking part in our study. 40 women initially participated in our rating study. Though, as one woman was already familiar with the study, she was excluded from analysis. Due to technical problems nine further participants had to be excluded from analysis as they were only able to rate less than half of the targets leaving us with a final sample size of 30 women for final analysis (mean age = 21.80 years, SD = 2.70, range = 19 - 29) . From these datasets, 26 were completed while four datasets only contained at least two thirds of all ratings, again due to technical issues. 13% of those women stated to have a university degree and 87% reported having a high school degree as their highest level of education. 18 (60%) women were currently involved in a romantic relationship. One woman reported being bisexual, all other women reported being heterosexual.

Rating study 1.6. The recruitment of our second replication took place in January to February 2016 at the University of Leipzig. We recruited 59 female raters (mean age = 21.29 years, SD = 3.54, range = 18 – 36). 32 (54%) of those women were currently involved in a romantic relationship, with 88% having a high school degree and 12% a university degree as their highest level of education. 49 (83%) women reported a heterosexual orientation and 10 (17%) women a bisexual orientation. Nearly all women except four women were fluent in German.

Table S3.

Number of ratings, mean values, standard deviations and ranges in each rating study as well as Cronbach's α and interclass correlation coefficients (ICC) for single ratings.

Variable	single ratings							aggregated ratings				
	N	*M*	*SD*	Min	Max	*ICC*	*α*	*n*	*M*	*SD*	Min	Max
	Rating Study 1.1											
Physical Attractiveness	1,657	2.82	1.53	1	7	.24	.93	88	2.82	0.79	1.42	5
	Rating Study 1.2											
Perceived intelligence	1,368	3.37	0.87	1	5	.22	.91	88	3.37	0.45	2.40	4.38
Perceived funniness	1,368	2.95	1.07	1	5	.28	.93	88	2.95	0.65	1.31	4.67
	Rating Study 1.3											
Short-term mate attractiveness	2,620	2.41	1.68	1	7	.36	.96	88	2.41	1.03	1.03	5.03
Long-term mate attractiveness	2,618	2.64	1.68	1	7	.27	.94	88	2.64	0.90	1.17	4.53
	Rating Study 1.4 - 1.6											
Short-term mate attractiveness	33,358	-.01	2.25	-4.22	8.36	n/a	n/a					
Long-term mate attractiveness	33,358	-.01	2.18	-4.13	8.40	n/a	n/a					

Note. n/a = not applicable. *ICC2* = Intraclass correlation coefficients for a random set of judges who rate every target. α = Cronbach's alpha.

C. Further information on Bayesian models

Our Bayesian models were fitted in Stan (Carpenter et al. 2016) via brms (version 2.10.0, Bürkner, 2017). Because of the varying rating scales across studies and the possibility that raters would use scales differently, we used mixed effects location-scale models, which allowed not only the mean response but also the residual variance to differ by study, Cue, target, and rater. These more complex models fit better according to the approximative leave-one-out information criterion (LOO-IC), although the main results did not change. To improve convergence and sampling efficiency, we used weakly informative priors, specifically normal(M=0,SD=5) for the non-varying effects on the means, Cauchy(0,3) for the varying effects on the means, N(0,1) for non-varying and varying effects on the residual variation. We fit four parallel chains to assess convergence using the Rhat statistic. The full code of the models is documented in our repository on OSF.

S3. Robustness checks and additional analyses

A. Hypothesis 1: Single perceiver accuracy

To investigate whether women can accurately perceive men's intelligence, we correlated men's *g* factor with an aggregated value of their perceived intelligence. Though as aggregated values tend to inflate accuracy estimates, we also used disaggregated ratings to determine the accuracy of individual women's judgements of intelligence (β = .18, p <.001, 95% CI [.07; .28]) in a structural equation model, modelling *g* as a hierarchical latent variable to correct for measurement error (see Table S4). This model fit our data well χ^2 (24, 2581 observations clustered in 88 targets) = 421.47 , p < .001 (comparative fit index [CFI] = .919, normative fit index [NFI], standardized root mean square residual [SRMR] = .038, root mean squared error of approximation [RMSEA] = .08, 90% CI [.073, .087]). The results from both methods support our first predictions, suggesting that women are able to perceive intelligence with some degree of accuracy based on our three stimulus types.

Table S4.

Results of structure equation model on the accuracy of intelligence perception where the g factor is modelled as a hierarchical latent variable to correct for measurement error.

Term	Estimate	*p*	95 % CI
		Latent variables	
Reaction Time			
CRT	1.15	.40	[-1.51, 3.80]
SRT	.34	.36	[-.39, 1.06]
Language Tests			
BEFKI	.50	.29	[-.43, 1.43]
MWT	1.80	.31	[-1.62, 5.3]
ICAR			
Verbal	.65	<.001	[.33, .98]
Letter	1.04	<.001	[.51, 1.56]
Rotation	1.01	<.001	[.48, 1.55]
matrices	.72	<.001	[.32, 1.11]
g Factor			
ICAR	.88	.05	[-.01, 1.77]
Reaction Time	.18	.50	[-.34, .69]
Language	1.78	.41	[-2.47, 6.03]
		Regressions	
g ~ Intelligence response	.22	.005	[.07, .38]

B. Hypothesis 2 to 4: Results on LTM Attractiveness

As described, women's ratings of men's long-term mate and short-term mate attractiveness was highly correlated. Hence, results were extremely similar. We therefore decided to report only results on short-term mate attractiveness in our main manuscript and report results on long-term mate attractiveness as part of our supplement.

We assumed that more intelligent men were rated as more attractive as a potential partner. For short-term mate attractiveness, we found a contradicting effect (Table S10): more intelligent men were rated as less attractive as a short-term mate (b = -.07, 95% HDI [-.11; -.03]). For long-term mate attractiveness, we found no association between *g* and long-term mate attractiveness (b = -.02, 95% HDI [-.06; .01]).

With the previous results being based on female raters participating in study 1.4 - 1.6, we found similar results in a second set of raters. In study 1.3 women also rated target's short-term and long-term mate attractiveness. The difference is that women saw cue 4 (pantomime) instead of cue 3 (vowels) and rated men's mate appeal only once after watching all three videos. In this second set of female raters, we investigated our second hypothesis namely whether more intelligent men had a higher mate appeal once using aggregated (Table S5) and second using single ratings (Table S6). In a subsequent analyses we included the perception of men's intelligence in our model (Table S7).

In this second set of raters we replicated results of our main model: more intelligent men did not have a higher mate appeal. When analysing single ratings and not aggregated ratings, more intelligent men were even rated as less attractive as a short-term mate. In sum, we found no support for our hypothesis stating that more intelligent men have a higher mate appeal. Only men who were perceived to be more intelligent and men who were physically more attractive had a higher mate appeal.

Table S5.

Results of linear model predicting aggregated short-term and long-term mate attractiveness

	Short-term mate attractiveness			Long-term mate attractiveness		
Term	Estimate	*p*	95 % CI	Estimate	*p*	95% CI
Intercept	-.73	.002	[-1.18; -.28]	.05	.818	[-.42; .53]
g factor	.01	.902	[-.11; .13]	.10	.121	[-.03; .23]
physical attractiveness	1.11	<.001	[.96; 1.27]	.92	<.001	[.76; 1.08]

Note. 88 Targets were rated by n = 30 women rating men's short-term/long-term mate attractiveness and n = 19 women rating men's physical attractiveness.

Table S6.

Results of linear model predicting single ratings of short-term and long-term mate attractiveness

	Short-term mate attractiveness			Long-term mate attractiveness		
Term	Estimate	*p*	95 % CI	Estimate	*p*	95% CI
Intercept	1.62	<.001	[1.28; 1.96]	1.99	<.001	[1.64; 2.34]
g factor	-.16	=.038	[-.32; -.01]	-.04	.581	[-.19; .10]
physical attractiveness	.28	<.001	[.20; .36]	.23	<.001	[.16; .30]
Observations	*n* = 49,316 ratings			*n* = 49,316 ratings		

Note. 88 target men were rated by *n* = 30 raters of short-term/long-term mate attractiveness and *n* = 19 raters of physical attractiveness.

Table S7

Results of linear model predicting single ratings of short-term and long-term mate attractiveness with perceived intelligence.

	Short-term mate attractiveness			Long-term mate attractiveness		
Term	Estimate	*p*	95 % CI	Estimate	*p*	95% CI
Intercept	1.04	<.001	[.60; 1.48]	1.12	<.001	[.73; 1.51]
g factor	-.19	=.013	[-.34; -.04]	-.08	=.210	[-.21; .05]
Perceived intelligence	.17	=.002	[.06; .29]	.26	<.001	[.17; .36]
physical attractiveness	.27	<.001	[.20; .35]	.22	<.001	[.16; .29]
Observations	*n* = 767,989 ratings			*n* = 767,400 ratings		

Note. 88 targets were rated by *n* = 30 raters of short-term/long-term mate attractiveness, *n* = 19 raters of physical attractiveness and *n* = 16 raters on perceived intelligence.

We assumed that above the influence of intelligence, funnier men have a higher mate appeal. We predicted men's short-term and long-term mate attractiveness, respectively, with men's *g* factor and ratings of their funniness. Replicating our previous results, men's actual intelligence did not influence their mate appeal. Again, when analysing single ratings more intelligent men were rated as less attractive as a short-term mate. However, men who were perceived to be funnier had a higher mate appeal (Table S8, Table S9). Despite the significant effect of funniness, results do not support our prediction. Perceived funniness contributes to mate appeal independently of men's intelligence. Hence, funniness does not seem to be an indicator of intelligence.

Table S8.

Results of linear model predicting aggregated short-term and long-term mate attractiveness with funniness

	Short-term mate attractiveness			Long-term mate attractiveness		
Term	Estimate	p	95 % CI	Estimate	*p*	95% CI
Intercept	-1.59	<.001	[-2.07; 1.10]	-1.02	<.001	[-1.48; -.57]
g factor	.00	.938	[-.10; 0.11]	.10	.056	[.00; .19]
Perceived funniness	.55	<.001	[.36; .74]	.69	<.001	[.51; .87]
Physical attractiveness	.84	<.001	[0.68; 1.00]	.58	<.001	[.42; .73]

Note. 88 Targets were rated by $n = 30$ women rating men's short-term/long-term mate attractiveness, $n = 16$ women rated men's funniness and $n = 19$ women rating men's physical attractiveness.

Table S9

Results of linear model predicting single ratings of short-term and long-term mate attractiveness with funniness.

	Short-term mate attractiveness			Long-term mate attractiveness		
Term	Estimate	p	95 % CI	Estimate	p	95% CI
Intercept	.69	<.001	[.35; 1.03]	1.06	<.001	[.68; 1.44]
g factor	-.14	.034	[-.26; -.01]	-.02	.792	[-.13; .10]
Perceived funniness	.35	<.001	[.26; .45]	.35	<.001	[.26; .44]
Physical attractiveness	.24	<.001	[.17; .30]	.19	<.001	[.13; .24]
Observations	$n = 767{,}989$ ratings			$n = 767{,}400$ ratings		

Note. 88 target men were rated by $n = 30$ raters of short-term/long-term mate attractiveness, $n = 19$ raters of physical attractiveness and $n = 16$ raters of funniness.

And lastly, we assumed that when shifting from only physical attractiveness information (cue 1 -3), to additional cues related to men's intelligence (cue 4), short-term mate and long-term mate attractiveness of more intelligent men would increase. We predicted a further increase for more intelligent men after presenting additional intelligence information (cue 6: make experimenter laugh).

Similar to our results on short-term mate attractiveness, we found none of the predicted effects for men's long-term mate attractiveness: attractiveness ratings for more intelligent men increased neither after cue 4 (*g* factor x Cue 4: $b = .02$, 95% HDI [-.01; .06]) nor after cue 6 (*g* factor x Cue 4: b = .04, 95% HDI [-.00; .08]).

Table S10

Estimated associations from a Bayesian mixed effect location scale model.

	Estimated effect on each outcome [95% HDI]			
	Short-term mate attractiveness		**Long-term mate attractiveness**	
Term	**Estimate**	**95% CI**	**Estimate**	**95% CI**
non-varying				
Intercept	+0.44	+0.02;+0.87	+0.44	-0.01;+0.87
σ Intercept	-0.03	-0.18;+0.13	-0.24	-0.43;-0.05
Study 1.4	-0.50	-1.04;+0.05	-0.59	-1.16;+0.00
Study 1.5	-0.71	-1.20;-0.19	-0.65	-1.16;-0.11
Cue 1-2	-0.30	-0.41;-0.20	-0.27	-0.37;-0.17
Cue 4	+0.16	+0.07;+0.24	+0.11	+0.04;+0.20
Cue 6	+0.36	+0.23;+0.50	+0.26	+0.13;+0.38
phys. attractiveness	+1.15	+1.05;+1.24	+1.01	+0.93;+1.09
g factor	-0.07	-0.11;-0.03	-0.02	-0.06;+0.01
Study 1.4:Cue 1-2	+0.26	+0.12;+0.39	+0.23	+0.09;+0.36
Study 1.5:Cue 1-2	+0.25	+0.11;+0.39	+0.21	+0.08;+0.35
Study 1.4:Cue 4	-0.05	-0.16;+0.04	-0.04	-0.14;+0.06
Study 1.5:Cue 4	-0.05	-0.16;+0.05	+0.00	-0.10;+0.11
Study 1.4:Cue 6	-0.19	-0.36;-0.03	-0.15	-0.31;+0.00
Study 1.5:Cue 6	-0.26	-0.43;-0.09	-0.17	-0.32;-0.02
Cue 1-2:phys. attractiveness	-0.10	-0.14;-0.06	-0.06	-0.10;-0.03
Cue 4:phys. attractiveness	+0.07	+0.04;+0.11	+0.06	+0.02;+0.10
Cue 6:phys. attractiveness	+0.14	+0.10;+0.19	+0.13	+0.09;+0.18
Cue 1-2:*g* factor	-0.01	-0.04;+0.02	-0.01	-0.04;+0.01
Cue 4:*g* factor	+0.01	-0.02;+0.04	+0.02	-0.01;+0.06
Cue 6:*g* factor	+0.02	-0.02;+0.06	+0.04	-0.00;+0.08
σ Study 1.4	+0.28	+0.08;+0.47	+0.55	+0.31;+0.81
σ Study 1.5	+0.29	+0.11;+0.47	+0.47	+0.24;+0.69
σ Cue 1-2	-0.11	-0.15;-0.08	-0.14	-0.17;-0.10
σ Cue 4	+0.13	+0.08;+0.17	+0.16	+0.12;+0.21
σ Cue 6	+0.29	+0.24;+0.33	+0.34	+0.29;+0.38
σ Study 1.4:Cue 1-2	+0.12	+0.07;+0.17	+0.14	+0.09;+0.19
σ Study 1.5:Cue 1-2	+0.09	+0.03;+0.14	+0.11	+0.06;+0.17
σ Study 1.4:Cue 4	-0.04	-0.10;+0.02	-0.07	-0.13;-0.01
σ Study 1.5:Cue 4	-0.04	-0.10;+0.01	-0.05	-0.11;+0.00
σ Study 1.4:Cue 6	-0.09	-0.15;-0.03	-0.14	-0.20;-0.08
σ Study 1.5:Cue 6	-0.05	-0.11;+0.01	-0.05	-0.11;+0.01
rater (n=114)				
sd(Intercept)	+1.13	+0.98;+1.30	+1.17	+1.02;+1.34
sd(phys. attractiveness)	+0.46	+0.40;+0.53	+0.44	+0.38;+0.51
sd(*g* factor)	+0.16	+0.13;+0.20	+0.15	+0.12;+0.18
sd(Cue 1-2)	+0.21	+0.17;+0.26	+0.21	+0.16;+0.26
sd(Cue 4)	+0.11	+0.07;+0.15	+0.10	+0.05;+0.15
sd(Cue 6)	+0.26	+0.21;+0.32	+0.24	+0.18;+0.30
sd(phys. attractiveness:Cue 1-2)	+0.06	+0.03;+0.09	+0.03	+0.01;+0.06
sd(phys. attractiveness:Cue 4)	+0.02	+0.00;+0.06	+0.03	+0.01;+0.06
sd(phys. attractiveness:Cue 6)	+0.07	+0.03;+0.11	+0.09	+0.05;+0.14
sd(*g* factor:Cue 1-2)	+0.02	+0.00;+0.05	+0.01	+0.00;+0.04
sd(*g* factor:Cue 4)	+0.01	+0.00;+0.04	+0.03	+0.00;+0.06
sd(*g* factor:Cue 6)	+0.02	+0.00;+0.06	+0.04	+0.00;+0.08
sd(σ Intercept)	+0.38	+0.33;+0.43	+0.48	+0.42;+0.55
target (n=88)				
sd(Intercept)	+0.19	+0.16;+0.24	+0.08	+0.06;+0.10
sd(Cue 1-2)	+0.07	+0.04;+0.11	+0.03	+0.01;+0.05
sd(Cue 4)	+0.08	+0.04;+0.12	+0.03	+0.00;+0.05

	Estimated effect on each outcome [95% HDI]			
	Short-term mate attractiveness		**Long-term mate attractiveness**	
Term	**Estimate**	**95% CI**	**Estimate**	**95% CI**
sd(Cue 6)	+0.14	+0.09;+0.19	+0.12	+0.08;+0.15
sd(σ Intercept)	+0.28	+0.24;+0.33	+0.28	+0.24;+0.33

Note:
Estimated associations from a Bayesian mixed effect location scale model. Estimates prefixed σ denote estimates on the residual standard deviation.

C. Robustness Checks: Inclusion of only heterosexual targets and raters

We pre-registered to recruit only heterosexual participants. However, three of our target men reported a bisexual or homosexual orientation. 12 of our female raters reported a bisexual or homosexual orientation. We excluded these 15 participants and reran our analyses. Replicating our results, more intelligent men were not preferred as a potential partner (for short-term mate attractiveness b = -.07, 95% HDI [-.11;-.03], for long-term mate attractiveness b = -.03, 95% HDI [-.06, .01]). Additionally, attractiveness did not increase after increasing information on men's intelligence (for short-term mate attractiveness: cue 4 * *g* factor b = .01, 95% HDI [-.02, .05], cue 6 * *g* factor b = .03, 95% HDI [-.01, .07]; for long-term mate attractiveness: cue 4 * *g* factor b = .03, 95% HDI [-.01, .07], cue 6 * g factor b = -.04, 95% HDI [-.00, .09]).

Table S11

Estimated associations from a Bayesian mixed effect location scale model including only heterosexual participants.

	Estimated effect on each outcome [95% HDI]			
	Short-term mate attractiveness		**Long-term mate attractiveness**	
Term	**Estimate**	**95% CI**	**Estimate**	**95% CI**
non-varying				
Intercept	+0.41	-0.02;+0.85	+0.42	-0.04;+0.89
σ Intercept	-0.04	-0.20;+0.13	-0.24	-0.46;-0.04
Study 1.4	-0.53	-1.10;+0.02	-0.61	-1.24;+0.01
Study 1.5	-0.66	-1.17;-0.14	-0.60	-1.17;-0.03
Cue 1-2	-0.30	-0.40;-0.19	-0.26	-0.37;-0.16
Cue 4	+0.16	+0.08;+0.24	+0.11	+0.04;+0.19
Cue 6	+0.36	+0.23;+0.49	+0.25	+0.12;+0.37
phys. attractiveness	+1.11	+1.02;+1.20	+0.98	+0.89;+1.06
g factor	-0.07	-0.11;-0.03	-0.03	-0.06;+0.01
Study 1.4:Cue 1-2	+0.27	+0.14;+0.41	+0.24	+0.09;+0.38
Study 1.5:Cue 1-2	+0.23	+0.09;+0.37	+0.20	+0.06;+0.34
Study 1.4:Cue 4	-0.06	-0.16;+0.04	-0.04	-0.15;+0.06
Study 1.5:Cue 4	-0.03	-0.14;+0.08	+0.03	-0.08;+0.13
Study 1.4:Cue 6	-0.20	-0.37;-0.04	-0.15	-0.31;+0.00
Study 1.5:Cue 6	-0.22	-0.38;-0.05	-0.13	-0.28;+0.03
Cue 1-2:phys. attractiveness	-0.10	-0.13;-0.06	-0.06	-0.10;-0.03
Cue 4:phys. attractiveness	+0.07	+0.04;+0.11	+0.06	+0.02;+0.10
Cue 6:phys. attractiveness	+0.15	+0.10;+0.19	+0.14	+0.09;+0.18
Cue 1-2:*g* factor	-0.01	-0.04;+0.02	-0.01	-0.04;+0.01
Cue 4:*g* factor	+0.01	-0.02;+0.05	+0.03	-0.01;+0.07
Cue 6:*g* factor	+0.03	-0.01;+0.07	+0.04	-0.00;+0.09
σ Study 1.4	+0.28	+0.06;+0.49	+0.54	+0.28;+0.81
σ Study 1.5	+0.30	+0.10;+0.48	+0.47	+0.23;+0.72
σ Cue 1-2	-0.11	-0.15;-0.08	-0.14	-0.17;-0.10
σ Cue 4	+0.13	+0.08;+0.17	+0.17	+0.12;+0.21
σ Cue 6	+0.29	+0.24;+0.33	+0.34	+0.29;+0.38
σ Study 1.4:Cue 1-2	+0.12	+0.07;+0.17	+0.14	+0.09;+0.20
σ Study 1.5:Cue 1-2	+0.09	+0.03;+0.15	+0.12	+0.06;+0.18
σ Study 1.4:Cue 4	-0.04	-0.10;+0.02	-0.07	-0.13;-0.01
σ Study 1.5:Cue 4	-0.04	-0.10;+0.02	-0.06	-0.12;+0.00
σ Study 1.4:Cue 6	-0.10	-0.16;-0.04	-0.15	-0.21;-0.09
σ Study 1.5:Cue 6	-0.06	-0.12;+0.01	-0.05	-0.11;+0.01

	Estimated effect on each outcome [95% HDI]			
	Short-term mate attractiveness		**Long-term mate attractiveness**	
Term	**Estimate**	**95% CI**	**Estimate**	**95% CI**
rater (n=105)				
sd(Intercept)	+1.16	+1.01;+1.34	+1.21	+1.05;+1.39
sd(phys. attractiveness)	+0.45	+0.39;+0.52	+0.43	+0.37;+0.50
sd(*g* factor)	+0.16	+0.12;+0.19	+0.14	+0.11;+0.18
sd(Cue 1-2)	+0.21	+0.17;+0.27	+0.21	+0.16;+0.27
sd(Cue 4)	+0.10	+0.07;+0.14	+0.09	+0.04;+0.14
sd(Cue 6)	+0.26	+0.21;+0.32	+0.23	+0.17;+0.30
sd(phys. attractiveness:Cue 1-2)	+0.06	+0.03;+0.10	+0.03	+0.01;+0.07
sd(phys. attractiveness:Cue 4)	+0.02	+0.00;+0.06	+0.03	+0.01;+0.06
sd(phys. attractiveness:Cue 6)	+0.07	+0.03;+0.11	+0.09	+0.04;+0.14
sd(*g* factor:Cue 1-2)	+0.02	+0.00;+0.05	+0.01	+0.00;+0.04
sd(*g* factor:Cue 4)	+0.01	+0.00;+0.04	+0.03	+0.00;+0.06
sd(*g* factor:Cue 6)	+0.02	+0.00;+0.06	+0.04	+0.00;+0.09
sd(σ Intercept)	+0.39	+0.34;+0.45	+0.50	+0.44;+0.58
target (n=88)				
sd(Intercept)	+0.19	+0.16;+0.23	+0.08	+0.05;+0.10
sd(Cue 1-2)	+0.07	+0.03;+0.10	+0.03	+0.00;+0.05
sd(Cue 4)	+0.07	+0.03;+0.11	+0.02	+0.00;+0.05
sd(Cue 6)	+0.13	+0.08;+0.19	+0.12	+0.08;+0.15
sd(σ Intercept)	+0.29	+0.25;+0.34	+0.29	+0.25;+0.34

Note:
Estimated associations from a Bayesian mixed effect location scale model. Estimates prefixed σ denote estimates on the residual standard deviation.

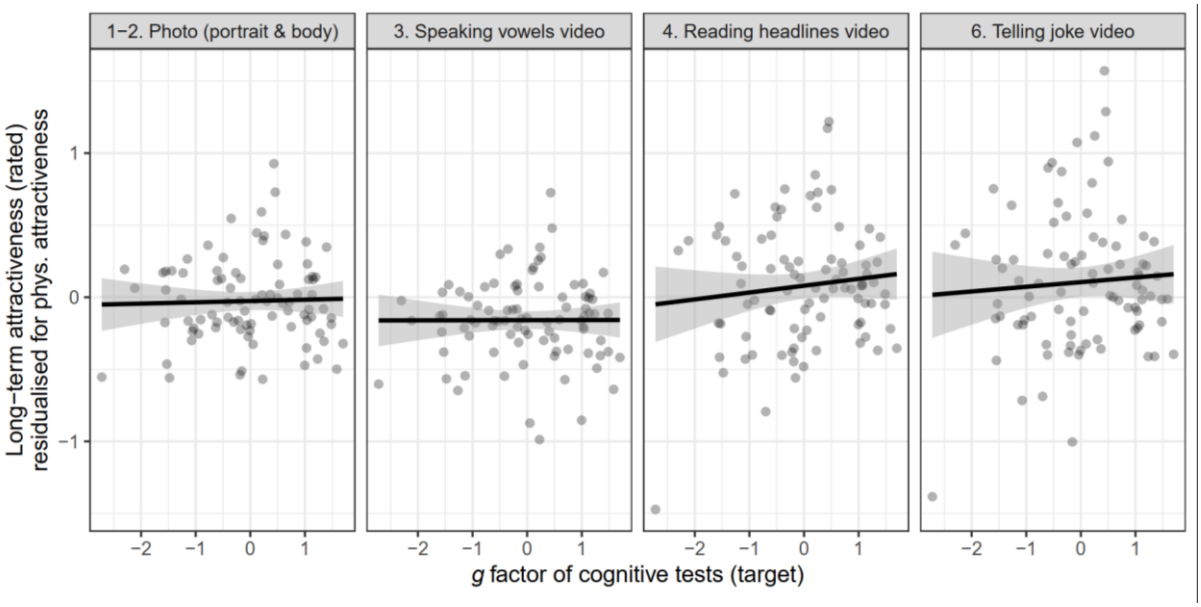
D. Robustness Check: No effects on sigma

Table S12.
Results for a simple mixed effects model

	Estimated effect on each outcome [95% CI]			
	Short-term mate attractiveness		**Long-term mate attractiveness**	
Term	**Estimate**	**95% CI**	**Estimate**	**95% CI**
non-varying				
Intercept	+0.41	-0.04;+0.86	+0.39	-0.08;+0.86
Study 1.4	-0.49	-1.08;+0.09	-0.53	-1.11;+0.09
Study 1.5	-0.60	-1.14;-0.07	-0.53	-1.07;+0.04
Cue 1-2	-0.38	-0.51;-0.24	-0.34	-0.46;-0.21
Cue 4	+0.12	+0.01;+0.23	+0.11	-0.01;+0.23
Cue 6	+0.33	+0.18;+0.49	+0.26	+0.09;+0.43
phys. attractiveness	+1.20	+1.11;+1.29	+1.06	+0.97;+1.15
g factor	-0.09	-0.15;-0.04	-0.05	-0.10;+0.00
Study 1.4:Cue 1-2	+0.32	+0.14;+0.49	+0.29	+0.12;+0.46
Study 1.5:Cue 1-2	+0.34	+0.17;+0.51	+0.29	+0.12;+0.45
Study 1.4:Cue 4	-0.03	-0.17;+0.12	-0.03	-0.18;+0.12
Study 1.5:Cue 4	+0.01	-0.13;+0.15	+0.02	-0.13;+0.17
Study 1.4:Cue 6	-0.18	-0.39;+0.01	-0.20	-0.41;+0.01
Study 1.5:Cue 6	-0.20	-0.39;-0.00	-0.14	-0.33;+0.05
Cue 1-2:phys. attractiveness	-0.14	-0.19;-0.09	-0.10	-0.15;-0.05
Cue 4:phys. attractiveness	+0.09	+0.04;+0.15	+0.06	+0.01;+0.12
Cue 6:phys. attractiveness	+0.18	+0.12;+0.23	+0.11	+0.05;+0.16
Cue 1-2:*g* factor	-0.01	-0.06;+0.04	-0.02	-0.07;+0.03
Cue 4:*g* factor	+0.04	-0.01;+0.09	+0.08	+0.03;+0.13
Cue 6:*g* factor	+0.05	-0.00;+0.11	+0.09	+0.04;+0.15
rater (n=114)				
sd(Intercept)	+1.12	+0.98;+1.30	+1.16	+1.01;+1.33
sd(phys. attractiveness)	+0.45	+0.39;+0.52	+0.44	+0.38;+0.51
sd(*g* factor)	+0.19	+0.16;+0.22	+0.17	+0.14;+0.21
sd(Cue 1-2)	+0.24	+0.18;+0.30	+0.24	+0.18;+0.30
sd(Cue 4)	+0.14	+0.08;+0.19	+0.16	+0.10;+0.21
sd(Cue 6)	+0.28	+0.22;+0.35	+0.30	+0.24;+0.37
sd(phys. attractiveness:Cue 1-2)	+0.07	+0.01;+0.12	+0.04	+0.00;+0.09
sd(phys. attractiveness:Cue 4)	+0.02	+0.00;+0.06	+0.02	+0.00;+0.06
sd(phys. attractiveness:Cue 6)	+0.04	+0.00;+0.10	+0.06	+0.00;+0.12
sd(*g* factor:Cue 1-2)	+0.02	+0.00;+0.06	+0.03	+0.00;+0.08
sd(*g* factor:Cue 4)	+0.02	+0.00;+0.06	+0.04	+0.00;+0.09
sd(*g* factor:Cue 6)	+0.03	+0.00;+0.09	+0.07	+0.01;+0.13
target (n=88)				
sd(Intercept)	+0.31	+0.26;+0.38	+0.26	+0.21;+0.31
sd(Cue 1-2)	+0.15	+0.10;+0.20	+0.13	+0.08;+0.19
sd(Cue 4)	+0.13	+0.08;+0.19	+0.19	+0.13;+0.25
sd(Cue 6)	+0.23	+0.17;+0.29	+0.28	+0.22;+0.35

Note:
Estimated associations from a Bayesian mixed effects model without allowing the residual variation to vary. In this model, the 95% HDI for the interactions between Cue 4/6 and the *g* factor on long-term mate attractiveness excluded zero, but this result was not robust in models that allowed residual variation to vary across rating studies.

Figure S1: The aggregated long-term mate attractiveness ratings made after seeing each cue (or set of cues) was adjusted for physical attractiveness. The plot shows the slope of a linear regression predicting short-term attractiveness from the measured *g* factor. More intelligent men were rated only slightly more favourably after intelligence-relevant information became available.



E. Results remain unchanged with known participants and partnered participants excluded

Table S13.
Study 2 results excluding known participants and partnered participants

	Rated Intelligence			**Rated Funniness**			**Rated Attractiveness**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Intercept	0.07	-0.01 – 0.16	0.083	-0.05	-0.14 – 0.03	0.212	-0.06	-0.15 – 0.03	0.209
Shipley (Vocabulary)	0.07	0.02 – 0.13	**0.006**	0.01	-0.05 – 0.06	0.846	-0.02	-0.07 – 0.04	0.608
Sex (Male)	-0.13	-0.25 – -0.01	**0.030**	0.13	0.01 – 0.25	**0.033**	0.15	0.02 – 0.27	**0.025**
N	671 $_{id}$			671 $_{id}$			671 $_{id}$		
	773 $_{partnerid}$			773 $_{partnerid}$			773 $_{partnerid}$		
Observations	1887			1887			1885		

	Rated Attractiveness			**Rated Funniness**			**Rated Attractiveness**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Intercept	-0.07	-0.16 – 0.01	0.093	-0.07	-0.15 – 0.01	0.080	-0.04	-0.12 – 0.04	0.355
Rated Intelligence	0.29	0.24 – 0.34	**<0.001**	0.29	0.24 – 0.34	**<0.001**			
Sex (Male)	0.18	0.06 – 0.30	**0.003**	0.16	0.04 – 0.27	**0.006**	0.09	-0.02 – 0.20	0.097
Rated Funniness							0.40	0.36 – 0.44	**<0.001**
N	672 $_{id}$			672 $_{id}$			672 $_{id}$		
	773 $_{partnerid}$			773 $_{partnerid}$			773 $_{partnerid}$		
Observations	1889			1891			1889		

F. No moderation by sex

Table S14.
Study 2 results including moderation by sex and all random effects.

	Rated Intelligence			**Rated Funniness**			**Rated Attractiveness**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Intercept	0.07	-0.01 – 0.15	0.070	-0.07	-0.15 – 0.02	0.119	-0.05	-0.14 – 0.03	0.234
Shipley (Vocabulary)	0.08	0.01 – 0.15	**0.028**	0.00	-0.07 – 0.07	0.990	-0.04	-0.12 – 0.04	0.295
Sex (Male)	-0.14	-0.25 – -0.03	**0.016**	0.14	0.02 – 0.25	**0.022**	0.13	0.01 – 0.26	**0.031**
Shipley (Vocabulary):Sex	0.00	-0.10 – 0.10	0.972	-0.01	-0.12 – 0.09	0.816	0.05	-0.06 – 0.16	0.389
Random Effects									
σ^2	0.61			0.56			0.47		
τ_{00}	0.20 $_{\text{partnerid}}$			0.20 $_{\text{partnerid}}$			0.20 $_{\text{partnerid}}$		
	0.17 $_{\text{id}}$			0.23 $_{\text{id}}$			0.32 $_{\text{id}}$		
τ_{11}	0.01 $_{\text{partnerid. Shipley (V):sex}}$			0.01 $_{\text{partnerid. Shipley (V):sex}}$			0.02 $_{\text{partnerid. Shipley (V):sex}}$		
ρ_{01}	-0.04 $_{\text{partnerid}}$			0.07 $_{\text{partnerid}}$			0.38 $_{\text{partnerid}}$		
ICC	0.38			0.44			0.53		
N	727 $_{\text{id}}$			727 $_{\text{id}}$			727 $_{\text{id}}$		
	786 $_{\text{partnerid}}$			786 $_{\text{partnerid}}$			786 $_{\text{partnerid}}$		
Observations	2116			2116			2114		

	Rated Attractiveness			Rated Funniness			Rated Attractiveness		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Intercept	-0.07	-0.15 – 0.01	0.087	-0.09	-0.16 – -0.01	**0.026**	-0.03	-0.10 – 0.05	0.478
Rated Intelligence	0.26	0.20 – 0.32	**<0.001**	0.26	0.19 – 0.32	**<0.001**			
Sex	0.18	0.07 – 0.30	**0.002**	0.18	0.07 – 0.29	**0.001**	0.08	-0.03 – 0.19	0.141
Rated Intelligence:Sex	0.06	-0.02 – 0.15	0.157	0.09	0.00 – 0.18	**0.048**			
Rated Funniness							0.43	0.38 – 0.49	**<0.001**
Rated Funniness:Sex							-0.06	-0.14 – 0.01	0.106
Random Effects									
σ^2	0.39			0.49			0.42		
τ_{00}	0.15 $_{\text{partnerid}}$			0.14 $_{\text{partnerid}}$			0.15 $_{\text{partnerid}}$		
	0.30 $_{\text{id}}$			0.22 $_{\text{id}}$			0.24 $_{\text{id}}$		
τ_{11}	0.06 $_{\text{partnerid.(R. Intelligence)}}$			0.02 $_{\text{partnerid.(R. Intelligence)}}$					
	0.01 $_{\text{id.scale.(R. Intelligence)}}$			0.04 $_{\text{id.scale.(R. Intelligence)}}$					
ρ_{01}	0.17 $_{\text{partnerid}}$			0.33 $_{\text{partnerid}}$					
	0.16 $_{\text{id}}$			0.22 $_{\text{id}}$					
ICC	0.57			0.46			0.47		
N	728 $_{\text{id}}$			728 $_{\text{id}}$			728 $_{\text{id}}$		
	786 $_{\text{partnerid}}$			786 $_{\text{partnerid}}$			786 $_{\text{partnerid}}$		
Observations	2118			2120			2118		

G. Pattern of results remains the same when facial and bodily attractiveness is controlled

Table S15.
Study 2 results with facial and bodily attractiveness variables included in all models

	Rated Intelligence			**Rated Funniness**			**Rated Overall Attractiveness**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.09	0.02 – 0.17	**0.017**	-0.04	-0.12 – 0.03	0.286	-0.00	-0.04 – 0.04	0.921
Shipley (Vocabulary)	0.08	0.04 – 0.13	**0.001**	0.00	-0.04 – 0.05	0.901	-0.00	-0.03 – 0.02	0.780
Facial Attractiveness	0.19	0.13 – 0.24	**<0.001**	0.26	0.21 – 0.32	**<0.001**	0.59	0.56 – 0.63	**<0.001**
Body Attractiveness	0.11	0.05 – 0.17	**<0.001**	0.15	0.09 – 0.20	**<0.001**	0.29	0.25 – 0.32	**<0.001**
Sex (Male)	-0.17	-0.28 – -0.07	**0.001**	0.08	-0.02 – 0.19	0.123	0.01	-0.04 – 0.07	0.699
N	727 $_{id}$			727 $_{id}$			727 $_{id}$		
	786 $_{partnerid}$			786 $_{partnerid}$			786 $_{partnerid}$		
Observations	2115			2115			2113		

	Rated Overall Attractiveness			**Rated Funniness**			**Rated Overall Attractiveness**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	-0.01	-0.05 – 0.03	0.511	-0.06	-0.13 – 0.01	0.091	0.00	-0.04 – 0.04	0.947
Rated Intelligence	0.09	0.06 – 0.11	**<0.001**	0.21	0.17 – 0.25	**<0.001**			
Facial Attractiveness	0.58	0.54 – 0.61	**<0.001**	0.22	0.17 – 0.27	**<0.001**	0.55	0.52 – 0.59	**<0.001**
Bodily Attractiveness	0.28	0.24 – 0.31	**<0.001**	0.13	0.07 – 0.18	**<0.001**	0.27	0.24 – 0.30	**<0.001**
Sex (Male)	0.03	-0.02 – 0.09	0.262	0.12	0.02 – 0.23	**0.018**	0.00	-0.05 – 0.06	0.919
Rated Funniness							0.14	0.12 – 0.17	**<0.001**
N	728 $_{id}$			728 $_{id}$			728 $_{id}$		
	786 $_{partnerid}$			786 $_{partnerid}$			786 $_{partnerid}$		
Observations	2116			2118			2116		