

Abstract

Individuals consistently report preferring humour in a romantic partner; but it is unclear why. The ‘fitness indicator hypothesis’ proposes that attraction to humour evolved because it is an indicator of genetic fitness. Studies testing predictions from this hypothesis, mostly based on stated preferences regarding a hypothetical ideal partner or on artificial tasks or scenarios, have so far yielded conflicting evidence. Here, we assessed a sample of 554 (291 women) participants’ stated preferences for various traits including humour production and receptiveness, and their revealed preferences for the same traits through speed dates (i.e., a naturalistic, face-to-face setting). Dates were surreptitiously audio-recorded for a subset of 350 participants (188 women), enabling additional assessment of revealed preferences based on an objective measure of humour in the form of laughter frequency. We tested the predictions that 1) humour is an attractive trait, and 2) men are more attracted to humour receptivity compared to women, and women are more attracted to humour production compared to men. Stated preferences from men and women largely replicated those found in the existing literature and are consistent with the fitness indicator hypothesis. Results from revealed preferences found a main effect of funniness on ratings of overall partner attractiveness, but there was no significant effect of laughter on attractiveness. Revealed preferences, using both funniness ratings and laughter, also found no main effect of humour receptivity on overall attractiveness. Finally, we observed no sex differences in the effects of humour production and humour receptivity, as measured by both funniness ratings and laughter, on ratings of overall attractiveness.

Keywords: Humor; Laughter; Evolution; Mating; Speed-dating

Laughter and ratings of funniness in speed-dating do not support the fitness indicator hypothesis of humour

Despite the presence of humour in virtually all human cultures, its evolutionary basis is not well understood (Fry, 1994; Provine, 2000). While various evolutionary theories of humour have been proposed (see Polimeni & Reiss, 2006, and Kozbelt, 2019, for an overview), one possibility is that humour is conducive to mating success and thereby subject to sexual selection. Indeed, individuals consistently report a preference for humour in a romantic partner (Bressler et al., 2006; Buss, 1988; Daniel et al., 1985; McGee & Shevlin, 2009). However, it remains unclear why, in an ultimate sense, humour is considered an attractive quality. Humour is enjoyable, of course – but there is nothing *inherently* enjoyable or attractive about the kinds of stimuli that we describe as funny. The question becomes, why did we evolve to find certain kinds of stimuli funny and to be attracted to funny individuals?

Miller (2000b) suggests that humour production (attempts to amuse, e.g. making a joke), when successful, is a signal of underlying genetic quality. The idea is that successful humour production requires mental performance (e.g. speed, intelligence, creativity), which in turn requires a high-functioning brain, which in turn requires a low load of genetic mutations. According to the ‘fitness indicator hypothesis’, sexual or romantic preference for funny individuals is therefore beneficial because offspring will inherit lower mutation loads and pass on their parents’ genes more effectively. According to this sexual selection hypothesis, both being funny and being attracted to funny people are evolutionarily favoured.

Predicted sex differences

According to parental investment theory (Trivers, 1972), women should be more interested in fitness indicators in a partner than men. This is because women, compared with men, have had considerably greater minimum time and energy costs associated with rearing

offspring (e.g. nine months labour, then a long period of breastfeeding), thus making the consequences of a poor choice in partner more costly in terms of fitness. Conversely, given their much lower minimum investment in reproduction (i.e. a single act of sexual intercourse), men should on average be less choosy and more interested in accessing receptive women, to maximise their mating success (Darwin, 1871, Trivers, 1972). Therefore, if Miller's fitness indicator hypothesis is correct, one would expect a difference, on average, in how men and women produce and respond to humour. Men, more so than women, should attempt to produce high-quality humour and value receptivity towards it in a potential mate, whereas women, more so than men, should value high quality humour production in a potential mate.

Previous studies

Previous studies that sought to investigate the fitness indicator hypothesis have primarily relied on captioning tasks (e.g. writing an amusing caption to a cartoon), retrospective reports (e.g. having participants recall how funny they or other people are), or stated preferences (Bressler & Balshine, 2006; Bressler et al., 2006; Greengross & Miller 2011, Hone et al., 2015; Martin & Sullivan, 2013; Weisfeld et al., 2011). While many of these studies found results consistent with the fitness indicator hypothesis, the measures are problematic. Captioning tasks have little in common with live interactions, and stated preferences assume that participants have sufficient insight into their own preferences. Additionally, both stated preferences and retrospective reports are likely prone to memory or social desirability biases. In practice, stated preferences often fail to predict individual's evaluations of potential partners (i.e. their revealed preferences; Eastwick et al., 2014). Therefore, preferences for humour production and humour receptivity found using captioning tasks, retrospective reports, and in particular, stated preferences, may not generalise to live interactions, which have characterised mate evaluation throughout our evolutionary history.

Speed-dating

Speed-dating paradigms, which involve participants having brief one-on-one interactions (e.g. 3 to 5 minutes) with multiple partners each followed by a measure of attraction or willingness to date, provide a more ecologically valid way to test the fitness indicator hypothesis in that they involve face-to-face interactions as did mate selection in our evolutionary past. To be sure, speed-dating does not fully capture the mating process in our environment of evolutionary adaptiveness, given that they are short, one-off interactions. But attraction in speed dates predicts later romantic interest and dating (Baxter et al., 2022), suggesting that they to some extent capture processes of natural mate choice. Importantly, speed-dating results reflect revealed as opposed to stated preferences. Other studies such as Wilbur and Campbell (2011) measured revealed preferences using online dating profiles, but not with interactions or a behavioural measure of humour in the form of laughter. Despite these advantages, only a single study testing the fitness indicator hypothesis has employed a genuine speed-dating paradigm; McFarland et al. (2013) found that neither laughter in response to an interaction partner's humour (*at-partner laughter*) nor laughter at oneself (*at-self laughter*) predicted a participant's sense of connection with their partner. However, McFarland et al. (2013) collapsed both measures of laughter across men and women, meaning they could not examine whether the sexes differ in the value they place upon humour production.

Hall (2015) employed a similar paradigm to McFarland et al., (2013) (the pretence, however, was a study of how first impressions are formed) while looking at possible differences between men and women. They found no significant sex differences in the relationship between at-partner laughter and dating interest. However, in Hall (2015) participants had only one interaction partner, making it impossible to isolate participant-versus interaction-related variance, and to assess if any effects generalised across multiple

partners. Additionally, the interactions were partially scripted, reducing the experiment's ecological validity. Further, neither Hall (2015) nor McFarland et al. (2013) collected ratings by participants of their partners' funniness or humour receptivity. While post-interaction ratings such as these can be subject to a halo effect (Nisbett & Wilson, 1977), testing for an association between ratings of humour and attraction would still have allowed for an additional evaluation of the fitness indicator hypothesis. As well as the studies not collecting ratings, neither study assessed humour receptivity (overt displays of amusement in response to another's attempt at humour e.g. laughing). Finally, in both studies, participants were made aware that they were being recorded and studied, due to pre-approval of recordings and/or worn audio recording devices. If assessment and recordings are salient during interactions, participants may be less likely to act naturally, reducing the generalisability of findings to interpersonal interactions outside of an experimental setting.

Present Study

Our study advances upon previous tests of the fitness indicator hypothesis of humour by using a behavioural measure of both humour production and humour receptivity (i.e. laughter) in conjunction with unscripted interactions. We also asked participants to rate both funniness and humour receptivity, used inconspicuous recording devices, and collected a larger sample (861 interactions leading to 1722 datapoints for ratings). First, we tested predictions from the fitness indicator hypothesis using stated preferences: that women report valuing humour production in a partner more than men, and that men report valuing humour receptivity in a partner more than women. We then move on to testing predictions in a more ecologically valid setting by assessing revealed rather than stated preferences. We measure humour production and humour receptivity in a speed-date partner first by asking participants to rate these characteristics in each partner, as well as through measuring laughter frequency during the interactions. Preferences are revealed by how these measures are associated with

ratings of overall attractiveness. As well as testing whether these preferences for humour production and receptivity exist overall, we also test for the predicted sex differences: we expect that women would have a stronger revealed preference for humour production compared to men, while men would have a stronger revealed preference for humour receptivity compared to women.

Methods

Participants

Participants were 554 (291 women, 263 men) students from the University of Queensland. Ages of the participants ranged from 17 to 38 years ($M = 19.40$ years, $SD = 2.59$ years). Laughter during interactions was coded for a subset of 350 participants (188 women, 162 men), whose ages ranged from 17 to 36 years ($M = 19.45$ years, $SD = 2.58$ years). Participants were recruited through a research participation scheme from an introductory psychology course and in some cases through word of mouth. Course credit was awarded for participating. Eligibility criteria were English as a first language, being single, and being heterosexual.

Measures

Demographics. Items pertaining to demographic information such as age and sex were included.

Stated Preferences. Participants were asked to indicate their *preference* for both humour production (i.e. ‘How funny would you like them to be?’) and humour receptivity (i.e. How funny would you like them to find you?’) in an ideal opposite-sex romantic partner. Responses were recorded on a 7-point rating scale (1 = *Well Below Average*; 4 = *Average*; 7 = *Well Above Average*).

Participants were also asked to rate the *importance* of humour production (i.e. ‘They are funny’) and humour receptivity (i.e. ‘They find you funny’) in an ideal partner. Responses were recorded on a 7-point rating scale (1 = *Not at all*; 4 = *Somewhat Important*; 7 = *Very Important*).

Speed-dating ratings. Participants were asked to rate their partner’s overall attractiveness (i.e. ‘Overall, I would rate their attractiveness as...’), funniness (i.e. ‘They are funny’), and humour receptivity (i.e. ‘Based on the interaction you just had, how funny would they rate you?’). All responses were recorded on a 7-point rating scale (1 = *Well Below Average*; 4 = *Average*; 7 = *Well Above Average*).

Laughter. Humour receptiveness was operationalised as instances of a participant laughing either at their partner or themselves. This was measured using audio recordings.

An instance of laughter was defined as any audible vocalised exhaustion of air that would be characterised as laughter under normal circumstances (Bachorowski, 2001; Grammer, 1990). A continuous vocalisation of laughter was considered one instance. If a break occurred between vocalisations, then this was considered two distinct instances. Laughter was coded into two categories, 1) a participant laughing in response to their partner (*at-partner laughter*), and 2) a participant laughing at themselves (*at-self laughter*). Instances of at-partner laughter and at-self laughter were determined by who was last speaking. At-partner laughter occurred during or immediately after the interaction partner spoke, and at-self laughter occurred during or immediately after the participant themselves spoke. However, coders were instructed to exercise appropriate judgement depending on the context. For instance, if the participant laughed at their own speech, but this speech was a short statement of agreement with their interaction partner, then this was then coded as at-partner

laughter. If it was ambiguous who the participant was laughing at, this was coded as at-self laughter, as at-partner laughter is the variable of theoretical interest.

The first author coded instances of laughter for all available recorded interactions from both 2018 and 2019, a total of 569 interactions. Interactions were coded twice (with exceptions due to 85 misplaced files), once from the perspective of the participant, and once from the perspective of their partner. This resulted in a total of 1053 unique codings of laughter data coded by the first author. To assess the reliability of these ratings, a large subset of the data was coded by others: a second independent coder coded laughter for 87% of the 403 interactions from 2018, while a third independent coder coded laughter for all of the 165 interactions from 2019. Intraclass correlation coefficients (ICC) between raters for instances of laughter in each interaction were calculated for both years of interactions using the ‘irr’ package (Gamer et al., 2019) in version 4.1.1 of ‘R’ (R Core Team, 2021), based on absolute-agreement and a 2-way random-effects model. For the 2018 interactions, strong agreement was found for at-partner laughter, $ICC = .83$ (95% confidence intervals .78 to .86). At-self laughter also had a strong intraclass correlation coefficient, $ICC = .77$ (95% confidence intervals .74 to .80). For the 2019 interactions, strong agreement was found for at-partner laughter, $ICC = .91$ (95% confidence intervals .88 to .93). At-self laughter also had a strong intraclass correlation coefficient, $ICC = .90$ (95% confidence intervals from .87 to .92).

Procedure

Pre-speed-date. Details of the speed-meeting study were available prior to sign up on the UQ psychology research participation website. Information pertained to the nature of the interactions, eligibility criteria, and the possibility of being recorded. While participants were informed that participation was an opportunity to meet fellow psychology students in a speed-dating context, the current study never served as genuine speed-dating event where

subsequent dates with other participants could be formally organised, nor were participants ever led to believe such at any point during participation.

Before starting, participants were given an information sheet which described the procedure of the study, and communicated that participation was fully voluntary and that withdrawal from the study at any time was permitted without consequence. It also detailed the confidentiality of the data collected from the study and the possibility of audio recordings being taken. Once the study commenced, participants received an identification number and an Apple iPad that they used throughout the whole study. Each iPad was recording audio throughout the session. However, participants were only made aware of this until after the study (the iPad's visual notification of ongoing audio recording, a red light, was covered). This was done to avoid the knowledge of recordings influencing behaviour during interactions.

The laboratory setup consisted of four interaction stations, separated by 1.7m room dividers or walls, each with two chairs facing inward towards a table. Men and women initially sat in different adjoining rooms to avoid any inter-sex encounters before the interactions. During this stage, demographic information was collected using questionnaires and the procedure of the study was explained. A male and female participant were then directed to each interaction station. When participants had no interaction partner due to an uneven sex ratio, they were instructed to sit quietly for that round awaiting further instruction.

Speed-date. Interactions lasted three minutes, with a bell signalling the beginning and end of each one. Interactions consisted of unscripted conversations, with a maximum of four interactions occurring in a session, depending on attendance. After each interaction participants were instructed to quietly complete their ratings of their interaction partner while making sure that their iPad was not visible to their partner, so as not to reveal their ratings.

Once all participants finished their partner ratings, men or women were rotated for the next interaction until all opposite-sex dyads had interacted. Rotation was counterbalanced across sessions.

Post-speed-date. After their final interaction, participants completed their remaining self-report questionnaires. After every participant had completed these questionnaires, they were taken outside and given a verbal and written debrief. The verbal debrief highlighted that audio of each participant was recorded throughout the entire experiment. Participants were then informed that they could have their recordings deleted now, or at a later date, without consequence. It was reiterated that their data would remain anonymous. Of the total participants, only one asked for their audio to be deleted (this participant was excluded from all analyses). Once the session had ended, audio recordings which were not deleted were saved to the iPad and later edited and clipped to include only the three-minute interactions of the participant, for ease of coding. The audio editing software used was “Audacity” version 2.2.2 (Audacity Team, 2018).

Analysis

The design of this experiment, which involved participants giving ratings to multiple partners, lends itself to a hierarchical data structure. This means, for example, that ratings of opposite-sex partners are nested within each participant and their measured traits. As such, multi-level modelling (MLM) analyses were conducted (Kenny, Kashy, & Cook, 2006) using the ‘lme4’ and ‘lmerTest’ packages (Bates et al., 2015; Kuznetsova et al., 2017) in ‘R’ (R Core Team, 2021). All predictors were z-standardised at the appropriate level (i.e., interaction-level and participant-levels variables was done at interaction and participants levels, respectively). Random intercepts for the speed-dating session were included to account for variance attributable to factors such as the time of day and different

experimenters conducting the study. Random intercepts were also included for participants and partners due to the dyadic nature of the data; this accounts for both individual and partner differences. We also specified random slopes maximally so as not to inflate the Type 1 Error rate (Barr, 2013; Barr et al., 2013). Here, we report the estimated fixed effects for each model; for additional results, see the Supplementary Materials. Finally, the data and materials associated with this researcher have been made available in the Supplementary Materials.

Results

Stated preferences and importance of preferences for humour production and receptivity

Table 1.

Means and standard deviations of stated preferences and importance of preferences for both men and women (N = 554), as well as results for independent samples t-tests testing for differences between men and women.

Variables	Mean (SD)					
	Women	Men	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Preference for funniness	5.88 (0.85)	5.57 (0.98)	4.14	518.64	< .001	0.35
Preference that they find you funny	5.63 (0.99)	5.92 (0.91)	-3.75	551.88	< .001	-0.32
Importance of funniness	5.74 (1.01)	5.25 (1.22)	5.21	510.95	< .001	0.45
Importance that they find you funny	5.22 (1.21)	5.38 (1.19)	-1.39	544.57	.165	-0.12

Note. Preferences were recorded on a scale of 1 to 7 (1 = *Well Below Average*; 4 = *Average*; 7 = *Well Above Average*). Responses for importance were recorded on a scale of 1 to 7 (1 = *Not at all*; 4 = *Somewhat Important*; 7 = *Very Important*).

First, we tested for replication of the pattern of stated preferences found in earlier studies by testing for sex differences in humour production and receptivity preferences and importance. Results from these tests are displayed in Table 1. As predicted, women, compared to men, showed greater preference for funny partners and this preference was

regarded as more important. Also as predicted, men's preference for humour receptivity was greater than women's, but the predicted sex difference in the importance of this preference was not significant.

Revealed preferences for humour production and receptivity

MLM was used to evaluate the effects of 1) ratings of funniness and humour receptivity on overall attraction, 2) at-partner and at-self laughter on ratings of funniness, and 3) humour production and receptivity, as measured by laughter, on overall attraction. Means and standard deviations of relevant variables used in the MLM can be found in the Supplementary Materials.

Preliminary Tests

Before testing the key hypotheses, we performed two preliminary tests. First, we tested whether stated preferences for humour production and receptivity were associated with preferences revealed by the analyses of ratings of funniness or of laughter. To the contrary, we found that there was no positive interaction effect whereby stated preferences moderated associations between ratings of humour and partner evaluations, nor laughter and partner evaluations (see Supplementary Materials). This finding reinforces the importance of testing for sex differences in revealed preferences, since the stated preferences are not informative about revealed preferences for humour, and it is sex differences in the latter that we are primarily interested in. Second, we tested whether there was a significant association between at-self laughter and at-partner laughter. We found no significant association between at-self laughter and at-partner laughter, $\gamma = 0.01$, $p = .726$. Nonetheless, when testing the association between laughter and ratings of funniness, both at-partner laughter and at-self laughter were included in the model to estimate the unique contribution of each.

Association Between Ratings and Overall Attractiveness

Table 2.

MLM coefficients for the associations between ratings of funniness and overall attractiveness, ratings of humour receptivity and overall attractiveness, and interaction terms with sex (number of participants = 554).

Variables	Model-1 and Model-2				
	<i>Ratings of Overall Attractiveness</i>				
	γ	SE	<i>t</i>	<i>df</i>	<i>p</i>
Ratings of funniness	0.40***	0.03	12.42	107.22	< .001
Sex	-0.16*	0.07	-2.43	101.37	.017
Ratings of funniness x Sex	0.04	0.04	0.93	90.42	.355
Ratings of humour receptivity	0.03	0.03	1.00	424.41	.317
Sex	-0.22**	0.08	-2.86	89.88	.005
Ratings of humour receptivity x Sex	-0.01	0.05	-0.24	429.53	.807

*** $p < .001$; ** $p < .01$; * $p < .05$

Next, we moved on to testing our main hypotheses, first using partner ratings. As predicted, there was a significant main effect of ratings of funniness on ratings of overall attractiveness (Table 2). This means that participants who rated their partner as funnier judged their partner to be overall more attractive. Unexpectedly however, participants who rated their partner as more receptive to their humour did not rate their partner as overall more attractive. Further contrary to our predictions, neither the effect of ratings of funniness nor of humour receptivity differed significantly according to sex. This means that men and women placed similar value on humour production and on humour receptivity, as respectively measured by ratings of funniness and humour receptivity, in a speed-date partner.

Association Between Laughter and Ratings of Funniness

Table 3.

MLM coefficients for associations between at-partner laughter, at-self laughter, and sex, and ratings of funniness (number of participants = 350).

Variables	Model-3				
	<i>Ratings of Funniness</i>				
	γ	SE	<i>T</i>	<i>df</i>	<i>p</i>
At-partner laughter	0.18***	0.03	5.40	974.39	< .001
At-self laughter	0.08*	0.04	2.29	148.41	.024
Sex	-0.19*	0.08	-2.23	359.20	.026

*** $p < .001$; ** $p < .01$; * $p < .05$

We then sought to establish the validity of at-partner laughter as a measure of humour, while controlling for at-self laughter. As seen in Table 3, there was a significant main effect of at-partner laughter on ratings of funniness, such that participants who laughed more at their partner rated them funnier. This finding we expected, but we did not expect that there was also a significant (though much smaller) main effect of at-self laughter on ratings of funniness, such that participants who laughed more at themselves rated their partner as funnier. However, additional analyses revealed that including at-self laughter in models that tested whether at-partner laughter predicted overall attraction had no effect on the overall pattern of results (see Supplementary Materials). Therefore, original models not using at-self laughter are presented below.

Association Between Laughter and Overall Attractiveness

Table 4.

MLM coefficients for the associations between at-partner laughter and overall attractiveness, and at-partner laughter (received) and overall attractiveness, and interaction terms with sex (number of participants = 350).

Variables	Model-4 and Model-5				
	<i>Ratings of Overall Attractiveness</i>				
	γ	SE	<i>t</i>	<i>df</i>	<i>p</i>
At-partner laughter	0.06	0.03	1.73	116.38	.086
Sex	-0.24**	0.09	-2.68	137.33	.008
At-partner laughter x Sex	0.06	0.07	0.90	56.02	.375
At-partner laughter (received)	0.01	0.04	0.16	81.68	.856
Sex	-0.21*	0.09	-2.15	84.94	.032
At-partner laughter (received) x Sex	0.03	0.07	0.56	141.50	.630

*** $p < .001$; ** $p < .01$; * $p < .05$

We then tested our main hypotheses again, this time using laughter. Main effects of laughter, both given and received, were not significantly different from zero, and did not differ by sex, contrary to our predictions (Table 4). Specifically, there was no significant association between how much a participant laughed at their partner and their ratings of their partner's overall attractiveness, and participants who received more laughter from their partner did not rate them as more attractive overall. These associations did not differ significantly between men and women, meaning that women did not value humour production or receptivity, as measured by laughter, in a speed-date partner any more or less than men did.

We also re-ran our main analyses (Tables 2 and 4) while controlling for participant age, partner age, and the participant-partner age difference. The results of these analyses were consistent with the overall pattern of results of our original main analyses.

Discussion

We investigated the fitness indicator hypothesis for humour using a combination of stated and revealed preferences, the latter being based on both ratings (of partners' funniness and humour receptivity) and laughter during speed dates. Our findings showed a clear incongruence between the findings from stated and revealed preference measures: stated preferences showed a pattern of sex differences that largely supported the hypothesis, while revealed preferences showed no evidence at all for sex differences and thus do not fit easily with the hypothesis as it has been presented by Miller (2000a, 2000b).

Results from stated preferences were largely consistent with both previous findings (Bressler et al., 2006; Hone et al., 2015) and the sex differences predicted by the fitness indicator hypothesis. Women reported valuing funniness in an ideal partner more than men did, in terms of both the level of funniness preferred and the importance of that preference, and men desired a higher level of humour receptivity in an ideal partner than women did. However, because of growing questions about the validity of stated preferences, we aimed to test whether these sex differences were reflected in revealed preferences as well. Our first method for assessing revealed preferences was based on how strongly participants' ratings of their partners' funniness or humour receptivity (how funny does he/she think I am?) correlated with their ratings of the partners' overall attractiveness. This allowed us to first check a more basic premise of the fitness indicator hypothesis: that funniness is actually attractive. Consistent with this premise, partners who were rated as funnier were also rated as having greater overall attractiveness. Notably, the same was not found for humour receptivity

– partners rated as more receptive were rated no more or less attractive overall. More damaging for the fitness indicator hypothesis, though, is that the predicted sex differences were not observed at all: the associations of both funniness and humour receptivity with overall attractiveness were similar in men and women. These findings are consonant with those of Weisfeld et al. (2011) in a marriage context: higher ratings of humour production in a spouse were associated with greater marital satisfaction, but this effect did not significantly differ between the sexes.

There is good reason to think that revealed preferences as assessed by ratings are more valid than stated preferences (they are inferred from reported amusement at, and attraction to, real individuals as opposed to depending on insight and knowledge of one's own preferences), but they can still be problematic, as described earlier. For example, the correlations of ratings of funniness and overall attractiveness might be explained by the halo effect, which undermines their evidential strength in regard to the fitness indicator hypothesis. Therefore, to further investigate the possibility of humour as a fitness indicator, we examined revealed preferences using laughter as a behavioural measure of both humour production and receptivity. Previous studies (Hall, 2015; McFarland et al., 2013) that have used laughter to predict attraction did not test for an association between at-partner laughter and ratings of funniness. In our larger study, we found that at-partner laughter was significantly, positively associated with ratings of funniness. This effect suggests that at-partner laughter was, at least partially, a genuine reflection of participants' humour receptivity in response to their partner's humour production.

Using laughter as a measure of humour, we found evidence inconsistent with the fitness indicator hypothesis. Specifically, we observed no significant association between participants' at-partner laughter and their ratings of the partners' overall attractiveness. It is

worth noting the apparent contrast of this finding with the findings that at-partner laughter is significantly associated with ratings of funniness, and that ratings of funniness are significantly associated with ratings of overall attractiveness. One possibility is that the association between ratings of funniness and overall attractiveness may be due to the halo effect, whereas the 'true' component of funniness perception, captured to some extent by at-partner laughter as well as funniness ratings, is not associated with overall attractiveness. Another possible explanation is that there is an association between genuine laughter and overall attractiveness, but that not all laughter is a genuine response to humour production. For instance, a participant may laugh at something their partner said to reduce the awkwardness of the situation. This kind of awkward laughter would likely be negatively associated with overall attractiveness, which could help explain why we found no significant main effect of at-partner laughter on overall attractiveness. In any case, at face value our findings on the relationship between at-partner laughter and overall attractiveness are at odds with the predictions of the fitness indicator hypothesis.

In addition to measuring humour production behaviourally, we also employed a novel behavioural measure of humour receptivity by using laughter received by participants from their partners. We found, contrary to predictions, that laughter received by participants from their partner did not predict their ratings of their partner's overall attractiveness. This finding suggests that humour receptivity was not to be found attractive, contrary to previous research using stated preferences, which found that individuals report valuing humour receptivity in an opposite-sex partner (Bressler et al., 2006; Hone et al., 2015). It may be that people tend to be mistaken about their valuation of humour receptivity on stated preference items; or it may be that such items do not prompt participants to consider or report the downsides of over-receptivity, like excessive or indiscriminate laughing. In any case, while this finding is not helpful to the fitness indicator hypothesis, it is not necessarily a severe problem either, since

it only predicts that men value humour receptivity (i.e. an interaction effect), with no clear prediction regarding a main effect.

But our tests for sex differences in the attractiveness of humour receptivity, and in the attractiveness of humour production, provided no support for the fitness indicator hypothesis either. We found no evidence that women value humour production in a speed-date partner more than men do, or that men value humour receptivity in a speed-date partner more than women do. Again, these results for revealed preferences – using both partner ratings and laughter – contrast with our own and others’ findings for stated preferences, which is how most of the predicted sex differences were found (Bressler et al., 2006, Hone et al., 2015; Tornquist & Chiappe, 2015). Similar discrepancies have been found in other studies which have shown that previous demonstrations of sex differences in stated preferences (e.g. for physical attractiveness and earning potential) were not found in subsequent tests of revealed preferences (Eastwick et al., 2014). It may be that participants rely on cultural stereotypes (i.e. men produce humour, whereas women appreciate humour; Hitchens, 2007) when self-reporting their abstract preferences. However, during interactions and when subsequently rating their partners, such stereotypes are not needed because participants base their judgements off of real interactions. Recent evidence also suggests that failure to detect correspondence between stated and revealed preferences could simply be due to a lack of power to detect the relevant interaction when multiple traits inform mate evaluations (Zhao, et al., in preparation). In any event, given that revealed preferences are taken as a better reflection of one’s actual mating choices than are stated preferences, the results on sex differences from the current study, which do not support Miller’s (2000a, 2000b) fitness indicator hypothesis, might carry more weight than those using stated preferences which support the hypothesis.

We also re-ran our main analyses using participants' willingness to date their partner (i.e. Would you go on a date with this person? *Yes* or *no*) instead of overall attractiveness as the criterion (see Supplementary Materials). We found that results addressing the key predictions of the fitness indicator hypothesis (i.e. that funniness should be generally attractive, but that women should find funniness more attractive than men, whereas men should find humour receptivity more attractive than women) were consistent with those from our main analyses. However, inconsistent with the results of our main analyses, participants were more willing to date a partner who was more receptive to their humour. This was true whether measuring humour receptiveness with participants' ratings of humour receptivity or with laughter received by participants from their partners. A possible explanation for this inconsistency is that a participant being willing to go on a date with their partner may not require that they find their partner immediately attractive. A date may serve as an opportunity to get to know someone better before you decide whether one finds them attractive or not. If the only prerequisite is that that one's partner is agreeable and easy to get along with, for example, laughter in response to a participant's attempt at humour could make one more comfortable, and thus more willing to go on a date but not necessarily more attracted. In any event, while the main effects of humour receptivity, using both ratings and laughter, on willingness to date are consistent with the fitness indicator hypothesis, though the lack of a significant interaction with sex is not. Moreover, the overall pattern of results using willingness to date are generally inconsistent with the aforementioned key predictions of the fitness indicator hypothesis.

Overall, our results call into question not only the fitness indicator hypothesis, but also (or alternatively) the degree to which parental investment theory (Trivers, 1972) can be applied to sex differences in humans' preferences for fitness indicators. The absence of significant sex differences in revealed preferences alone does not necessarily exclude humour

as a fitness indicator; instead, it may be that theorists have overestimated the degree to which parental investment theory predicts sex differences in human fitness indicators. Humans exhibit mutual mate choice, where both males and females are choosy, whereas in other animals typically only females are choosy (Stewart-Williams & Thomas, 2013). Under mutual mate choice, fitness indicators are still expected to evolve, but not necessarily with large (or any) sex differences (Hooper & Miller, 2008). So the possibility remains that humour may be a fitness indicator, but men and women differ little, if at all, in their attraction to it. The current study is partially consistent with this possibility, as there is some evidence that humour production (per funniness ratings, but not laughter) and humour receptivity (per funniness rating and laughter) were desirable traits to both sexes similarly.

While the use of laughter as a behavioural measure of humour was a methodological strength of the current study, there are still problems with the current measure of laughter. As mentioned earlier, laughter by a participant, even if it is clearly at themselves or their partner, may not be genuine or even pertain to humour. A participant may laugh out of nervousness, to ease awkwardness, to be polite, or to appear happy. For instance, McFarland et al.'s (2013) speed-dating study used laughter as an indication of excitement with no consideration of it as a reflection of humour. While we found that at-partner laughter was a significant predictor of ratings of funniness, so too was at-self laughter (to a lesser extent). This suggests that there was some degree of extraneous variance in at-partner laughter as a measure of humour. Hall (2015) addressed this issue by coding for at-partner laughter only in response to attempts at humour production, but we decided against such an approach because of the greater subjectivity that would have been introduced by having to judge what is and what is not an attempt at humour production. Another issue is that there are various ways to code the laughter – coding of such a large volume of speed date recordings is laborious, and while we

aimed for the most sensible coding scheme, we cannot rule out variations in the results had we coded a different way.

A more general limitation of our methodology is that dates of a few minutes with strangers may not be optimal for observing the role of humour in courtship and attraction. For example, effective humour may depend on an understanding of the other person's sensibilities, which might be revealed only after longer acquaintance. We did observe variation in funniness ratings and laughter across dating partners, indicating these dates are neither devoid of, nor uniform in humour, but we would want convergent evidence from longer acquaintances to be more confident regarding the broader role of humour in attraction. A further consideration is that, though our sample is large, it may not be large enough to detect sex differences if they are very subtle, given that the sex differences in revealed preferences are tested by a three-way interaction (i.e. sex by humour measure by overall attractiveness rating).

In summary, we found that while stated preferences largely supported the sex differences predicted by the fitness indicator hypothesis, results from revealed preferences, which are taken as more valid than stated preferences, did not support these predicted sex differences and offered only mixed support for the central premise of the hypothesis, that funniness is attractive. It remains unclear how crucial the predicted sex differences are to the viability of the hypothesis, and more research is needed to analyse the possibility of different effects over longer periods of acquaintance.

References

- Audacity Team (2018). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.2.2 retrieved March 20th 2018 from <https://www.audacityteam.org/audacity-2-2-2-released/>
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4, 328. doi:10.3389/fpsyg.2013.00328
- Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Men Lang*, 68(3). doi: 10.1016/j.jml.2012.11.001
- Bachorowski, J.-A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *J Acoust Soc Am*, 110(3 I), 1581-1597. <https://doi.org/10.1121/1.1391244>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Baxter, A., et al. (2022). "Initial impressions of compatibility and mate value predict later dating and romantic interest." *Proc Natl Acad Sci U S A* 119(45): e2206925119.
- Blalock, H. M. (1969). *Theory construction : from verbal to mathematical formulations*. Prentice-Hall.
- Bressler, E. R., & Balshine, S. (2006). The influence of humor on desirability. *Evolution and human behavior*, 27(1), 29-39. <https://doi.org/10.1016/j.evolhumbehav.2005.06.002>
- Bressler, E. R., Martin, R. A., & Balshine, S. (2006). Production and appreciation of humor as sexually selected traits. *Evolution and human behavior*, 27(2), 121-130. <https://doi.org/10.1016/j.evolhumbehav.2005.09.001>
- Buss, D. M. (1988). The Evolution of Human Intrasexual Competition: Tactics of Mate Attraction. *Journal of personality and social psychology*, 54(4), 616-628. <https://doi.org/10.1037/0022-3514.54.4.616>
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.
- Davis, J. A. (1985). *The logic of causal order*. SAGE.
- Eastwick, P. W., Luchies, L. B., Finkel, E. J., & Hunt, L. L. (2014). The Predictive Validity of Ideal Partner Preferences: A Review and Meta-Analysis. *Psychol Bull*, 140(3), 623-665. <https://doi.org/10.1037/a0032432>
- Fry, W. F. (1994). The biology of humor. *Humor - International Journal of Humor Research*, 7(2), 111-126. <https://doi.org/10.1515/humr.1994.7.2.111>
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. <https://CRAN.R-project.org/package=irr>
- Grammer, K. (1990). Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters. *Journal of nonverbal behavior*, 14(4), 209-236. <https://doi.org/10.1007/BF00989317>
- Greengross, G., & Miller, G. (2011). Humor ability reveals intelligence, predicts mating success, and is higher in males. *Intelligence (Norwood)*, 39(4), 188-192. <https://doi.org/10.1016/j.intell.2011.03.006>
- Hal, J. D. (1985). Values in mate selection: A 1984 campus survey. *College Student Journal*, 19(1), 44-50.
- Hall, J. A. (2015). Sexual Selection and Humor in Courtship: A Case for Warmth and Extroversion. *Evolutionary psychology*, 13(3), 1-10. <https://doi.org/10.1177/1474704915598918>
- Hitchens, C. (2007). Why Women Aren't Funny. *Vanity fair (New York, N.Y.)*, 49(1), 54.

- Hone, L. S. E., Hurwitz, W., & Lieberman, D. (2015). Sex Differences in Preferences for Humor: A Replication, Modification, and Extension. *Evol Psychol*, 13(1), 167-181. <https://doi.org/10.1177/147470491501300110>
- Hooper, P. L., & Miller, G. F. (2008). Mutual Mate Choice Can Drive Costly Signaling Even Under Perfect Monogamy. *Adaptive behavior*, 16(1), 53-70. <https://doi.org/10.1177/1059712307087283>
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. Guilford Press.
- Kozbelt, A. (2019). Chapter 10 - Evolutionary Explanations for Humor and Creativity. *Creativity and Humor*. S. R. Luria, J. Baer and J. C. Kaufman, Academic Press: 205-230.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Li, N. P., Yong, J. C., Tsai, M. H., Lai, M. H. C., Lim, A. J. Y., & Ackerman, J. M. (2020). Confidence is sexy and it can be trained: Examining male social confidence in initial, opposite-sex interactions. *J Pers*, 88(6), 1235-1251. <https://doi.org/10.1111/jopy.12568>
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prev Sci*, 1(4), 173-181. <https://doi.org/10.1023/A:1026595011371>
- Martin, G. N., & Sullivan, E. (2013). Sense of humor across cultures: A comparison of British, Australian and American respondents. *North American journal of psychology*, 15(2), 375-384.
- McFarland, D. A., Jurafsky, D., & Rawlings, C. (2013). Making the connection: Social bonding in courtship situations. *The American journal of sociology*, 118(6), 1596-1649. <https://doi.org/10.1086/670240>
- McGee, E., & Shevlin, M. (2009). Effect of Humor on Interpersonal Attraction and Mate Selection. *J Psychol*, 143(1), 67-77. <https://doi.org/10.3200/JRLP.143.1.67-77>
- Miller, G. (2000). Mental Traits as Fitness Indicators: Expanding Evolutionary Psychology's Adaptationism. *Ann N Y Acad Sci*, 907(1), 62-74. <https://doi.org/10.1111/j.1749-6632.2000.tb06616.x>
- Miller, G. F. (2001). The mating mind: How sexual choice shaped the evolution of human nature. *Psychology (Washington, D.C.)*, 12(8), 1-21.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, 35(4), 250-256. <https://doi.org/10.1037//0022-3514.35.4.250>
- Polimeni, J. and J. P. Reiss (2006). "The First Joke: Exploring the Evolutionary Origins of Humor." *Evolutionary Psychology* 4(1): 147470490600400.
- Provine, R. R. (2000). *Laughter : a scientific investigation*. Penguin Books.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Stewart-Williams, S., & Thomas, A. G. (2013). The Ape That Thought It Was a Peacock: Does Evolutionary Psychology Exaggerate Human Sex Differences? *Psychological inquiry*, 24(3), 137-168. <https://doi.org/10.1080/1047840X.2013.804899>
- Tornquist, M., & Chiappe, D. (2015). Effects of Humor Production, Humor Receptivity, and Physical Attractiveness on Partner Desirability. *Evolutionary psychology*, 13(4), 147470491560874. <https://doi.org/10.1177/1474704915608744>
- Weisfeld, G. E., Nowak, N. T., Lucas, T., Weisfeld, C. C., Imamoğlu, E. O., Butovskaya, M., Shen, J., & Parkhill, M. R. (2011). Do women seek humorousness in men because it

signals intelligence? A cross-cultural test. *Humor - International Journal of Humor Research*, 24(4), 435-462. <https://doi.org/10.1515/humr.2011.025>

Wilbur, C. J. and L. Campbell (2011). "Humor in romantic contexts: do men participate and women evaluate?" *Pers Soc Psychol Bull* 37(7): 918-929.

Zhao, A. A. Z., Sidari, M. J., Lee, A. J., & Zietsch B. P. (in preparation). Using Speed-Dating and Simulation Data to Explain Inconsistent Findings on the Relation Between Stated and Revealed Preferences.