Supplemental Information: Humor ratings of bad jokes are modulated by other people's laughter for neurotypical and autistic adults.

Qing Cai, Sinead Chen, Sarah White & Sophie K. Scott

	Autism group	NT subgroup	NT group	Baseline group
N (male:female)	26 (21:5)	24 (18:6)	48 (28:20)	20 (8:12)
Age (years)	34.9 (7.7)	32.9 (10.9)	26.0 (7.38)	27.8 (4.011)
Verbal IQ	115.8 (10.1)	119.0 (10.0)	-	-
Performance IQ	110.5 (14.4)	115.5 (13.7)	-	-
AQ [a]	33.2 (8.7)	13.7 (6.0)	-	-
ADOS total	8.7 (3.4)	-	-	-
- communication subscale	3.3 (2.3)	-	-	-
- social subscale	5.7 (2.4)	-	-	-

Table S1 Background details of the participant groups
Note. Values are given as mean (standard deviation) except when otherwise stated. NT = Neurotypical; AQ = autism-spectrum quotient.
[a] one autistic participant did not complete the AQ questionnaire

Number	Jokes
1	What do you call a bear with no socks on? Bare foot.
2	What button is impossible to unbutton? The belly button.
3	What did Michael Jackson call his denim store? Billie Jeans.
4	What is invisible and smells of worms? A bird's fart.
5	What do you call a Minecraft celebration? A block party.
6	Who is the best Kung Fu vegetable? Brocc-Lee.
7	What do you call a rabbit who is angry over getting burnt? A hot cross bunny.
8	What did the butt say to the other butt? PTTTTT <raspberry noise>.
9	What's orange and sounds like a parrot? A carrot.
10	Why couldn't the toilet paper cross the road? He got stuck in a crack.
11	What's round and sounds like a trumpet? A crumpet.
12	What do you call a sleeping dinosaur? A dino-SNORE.
13	What do you call a man with a spade on his head? Dug.
14	What kind of hair do they sell at IHOP? Eggstensions.
15	What do you call an Asian man who always has correct change? Exact Lee.
16	What do you call an apple that farts? A fruity Tooty.
17	What is the best day to cook? FRY-DAY.
18	What did the horse say when it fell? GIDDYUP!
19	What do you receive when you ask a lemon to help? Lemon aid
20	Why can’t you give Elsa a balloon? Because she will let it go.
21	When does a sandwich cook? When it is bakin' lettuce and tomato.
22	Why did the smart phone need glasses? It lost all its contacts.
23	What state has the smallest drinks? Mini-soda.
24	Why did the cow cross the road? They wanted to go to the mooooovies.
25	What do you call a funky car? Mustang.
26	What did the hammer say to his homeboys? Nailed it.
27	Why are cats good at video games? Because they have nine lives.
28	What do you call a deer with no eyes? No idea.
29	What is big and green and falls off over the tree will kill you? A snooker table.
30	What did the French guy do when he drank too much water? He went oui oui in his pants
31	Where do pencils spend their vacation? Pencil-Vania.
32	What do flies eat for breakfast? A bowl of poop loops.
33	Why did the balloon go near the needle? He wanted to be a pop star.
34	What did the duck do when he read all these jokes? He quacked up.
35	Why was the tomato all red? It saw the salad dressing.
36	What do you call a female magician in the dessert? A sand witch.
37	How do billboards talk? Sign language.
38	What is brown and sticky? A stick.
39	I hurt my foot driving the other day. You know what I called? The toe company.
40	What does a dinosaur use to pay bills? Tyrannosaurus checks.

Table S2 Jokes used in the experiment

Jokes	Set 1	Set 2
J01	J01 + SpL	J01 + PoL
...	...	...
J20	J20 + SpL	J20 + PoL
J21	J21 + PoL	J21 + SpL
...	...	...
J40	J40 + PoL	J40 + SpL

Figure S1 How the two of jokes were paired with laughter stimuli (J=Joke, SpL=spontaneous laughter, PoL=posed laughter)

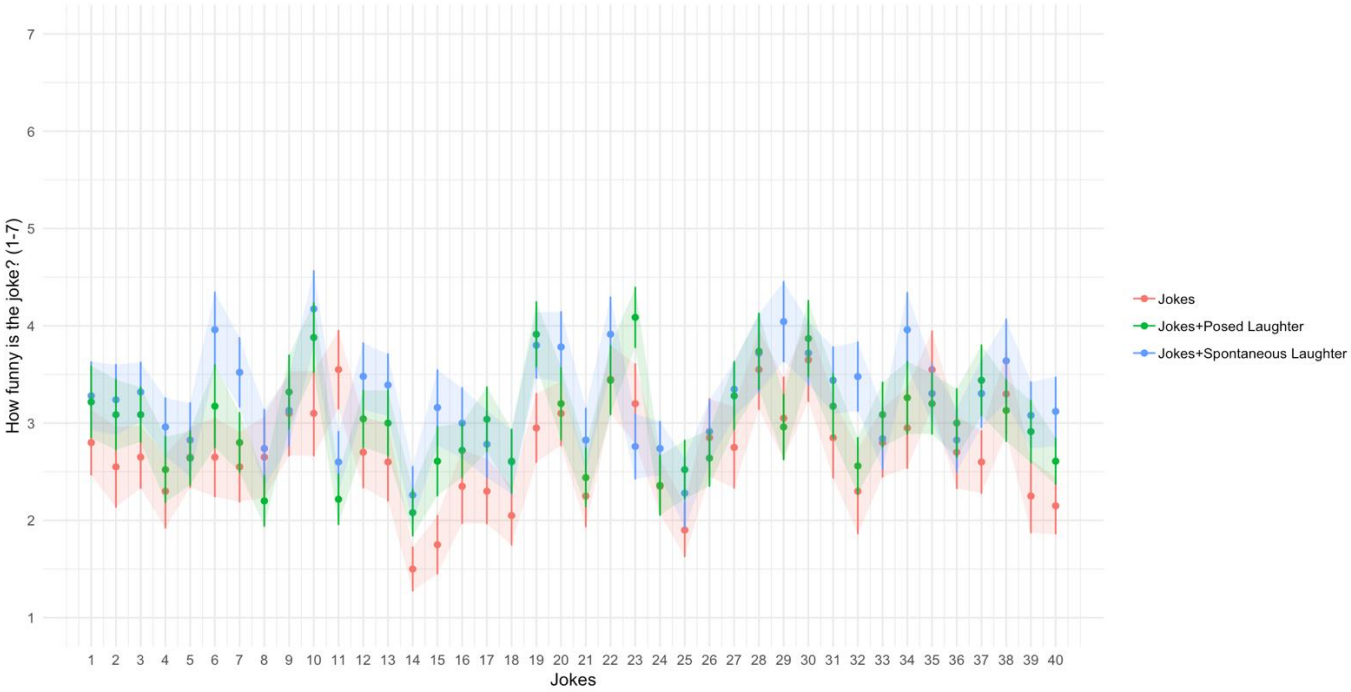

Figure S2 Average ratings for 40 jokes and 40 jokes paired either spontaneous or posed laughter in NT participants. Error bars represent standard error of mean.

Supplemental Experimental Procedures

Supplemental Participants

Participants were 26 autistic adults and 48 neuro-typical (NT) adults: this included a subgroup of 24 NT adults who were matched to the autism group for sex ($\chi^2(1) = .242$, $p = .623$), age ($t(48) = -.742$, $p = .462$), and verbal ($t(48) = 1.121$, $p = .268$) and performance ($t(48) = 1.234$, $p = .223$) IQ, as measured by the Wechsler Abbreviated Scale of Intelligence (WASI, [S1]) or Wechsler Adult Intelligence Scale (WAIS, [S2]). The groups differed in their self-report of autistic traits, measured by the Autism-Spectrum Quotient (AQ, [S3]), ($t(47) = -9.068$, $p < .001$). Full details of the groups are given in Table S1.

All participants in the autism group had a diagnosis of autism spectrum disorder (n = 5) or Asperger syndrome (n = 21) from a qualified clinician, with 12% reporting an additional diagnosis of another developmental disorders: dyslexia (n = 1), ADHD (n = 1) and dyspraxia

(n = 1). The Autism Diagnostic Observation Schedule (ADOS-G, [S4]) was administered to verify the diagnosis. In total, 9 participants met the criteria for autism and 11 more for autism spectrum on the ADOS classification. The remaining six scored below the threshold but were retained within the sample because five of them scored above the threshold for social symptoms on the ADOS, and they all reported significant difficulties in everyday life; this profile is frequently observed in autistic individuals with high IQ [S5]. Informed written consent was obtained prior to testing, and the project received approval from the UCL research ethics committee.

To establish baseline measures of the funniness (or otherwise) of the jokes we first recruited 20 college students, whose native language was English, to rate the original 40 jokes (these were not followed by the laughter stimuli). These ratings of the jokes also allowed us to generate two sets of jokes that were matched for their perceived amusement level. These two lists of jokes were used in the main experiment to be paired with spontaneous or posed laughter (see below).

Supplemental Materials

Laughter stimuli

The laughter stimuli (40 in total) consisted of 20 spontaneous (involuntary) and 20 posed (voluntary) laughter stimuli. We recorded spontaneous and posed laughter using the method previously validated in behavioural and neuroimaging experiments [S6, S7]. The laughter was generated by six adults (aged between 23 to 46 years; 3 females) who were not professional actors and was recorded using professional equipment in a sound-proof, anechoic chamber at University College London. Spontaneous laughter was elicited using an amusement induction situation in a social interactive setting: speakers were shown video clips, which they had identified beforehand as amusing and that would easily cause them to laugh aloud. The emotional experience was described positively by the adults during and after the recording session. For posed laughter, the adults were asked to simulate laughter in the absence of any external stimulation, and were encouraged to make it sound natural and positive. Importantly, speakers were always asked to produce posed laughter before the spontaneous laughter, to avoid the genuine emotional states associated with the latter affecting the production of posed laughter.

The raw audio files were downsampled at a rate of 44100 Hz to mono.wav files with 32-bit resolution. Individual files were prepared for each vocalisation from each adult by visually identifying the onset and offset of each event in their oscillograms. All files were then normalised for root-mean-square (RMS) amplitude using a speech analysis software called Praat [S8]. The original dataset (200 in total) contained 100 laughter stimuli (average duration = 2.51 seconds; SD = 0.36; range = 1.7 to 3.14 seconds). In order to select the best examples from the spontaneous and posed laughter dataset, a pilot perceptual validation experiment was conducted. Thirty native British speakers were asked to rate each stimulus on four different parameters (authenticity, emotion, frequency and control) using a 7-point Likert scales. Based on the results of the authenticity ratings (‘Does the sound reflected a genuinely-felt emotion?’, 1-signified posed, 7-signified genuine), 20 spontaneous laughter stimuli with the most highly-rated authenticity rating and 20 posed laughter which received the lowest authenticity rating were selected from the original dataset. Spontaneous and posed laughs were matched for duration (spontaneous laughs, M= 2530ms, SD = .385;

posed laughs, M= 2382ms, SD = .362), and the pilot validation data confirmed that the spontaneous laughs were perceived as highly authentic (spontaneous laughs, M = 5.80, SD = 0.45, on an authenticity scale from 1 to 7; posed laughs, M = 3.299, SD = 0.426).

Jokes paired with laughter stimuli

We recorded 40 jokes involving puns and wordplay, e.g. 'What is the best day to cook? FRY-DAY' as stimuli. Full details of the jokes are given in Table S2: we avoided jokes that rely on interpretation of intentions or social rules. The jokes were all somewhat puerile: this was to avoid a ceiling effect that might mask any effects of added laughter. The jokes were read aloud by a professional male comedian in a performance style. Recordings were made on a digital audio recorder. The raw audio files were downsampled at a rate of 44100 Hz to mono.wav files with 16-bit resolution. The duration of each joke stimulus was edited and cut into complete sound clips from 3 to 6 seconds. We further edited the jokes stimuli by randomly pairing each joke stimulus with either a spontaneous or a posed laughter stimulus. This process resulted in 80 jokes paired with laughter stimuli in total. Each combined stimulus was further edited into a separate .wav file (<9s each) using version 2.3.0 of Audacity(R) recording and editing software [S9].

Supplemental Experimental design and procedure

All possible pairings of joke and laughter type were counterbalanced across participants, such that we could determine the effects of laughter type on the perception of humour in the joke. To this end, the 80 jokes paired with laughter stimuli were assigned to two sets. Each set contained 40 stimuli with half of the jokes (J) paired with spontaneous laughter (SpL) and half paired with posed laughter (PoL); the second set contained the same 40 jokes but paired with the alternative form of laughter (Fig. S1). Participants listened to the 40 jokes paired with laughter stimuli from either Set 1 or Set 2. The stimuli were presented in a random order.

Testing was administered by one experimenter (Q.C.) in a testing room. All participants were asked to sit directly in front of the laptop and wear headphones. At the beginning of the study, participants were instructed that we had recorded some jokes read by a male comedian and different people's response to the jokes. To help this amateur comedian to get some feedback on his performance, the participants would listen to those recordings and make a judgment of how funny each recording was on a 7-point rating scale. Before the test session started, each participant confirmed that he or she understood the task and had had the opportunity to ask any questions.

The first trial began after participants clicked the start button. On each trial, participants were asked to rate the funniness of the joke stimulus on a 7-point Likert scale ('How funny is the joke?' 1 – Not funny at all, 7 - Extremely funny) presented on the screen as a slider that could be moved with the mouse; they had up to 6 seconds to respond. There were also eight catch trials in the experiment to ensure that participants paid attention to the stimuli, presented after every fifth joke stimulus. The catch trials required the participants to recall whether the laughter in the preceding joke stimulus was produced by a female or male speaker. Participants responded by clicking one of the figures on the screen representing a

female or male voice. The experiment took less than 10 minutes and was presented in Gorilla Experiment Builder [S10].

At the conclusion of testing, the experimenter asked whether participants had any further questions and gave them payment for taking part in the study. Participants were also encouraged to contact the researchers if they had any further questions.

Supplemental Results

All participants passed at least half of the catch trials, so we did not exclude any participant from the analysis.

Items were designed to produce a range of different degrees of funniness; we expected participants to rate some items much higher than others. For this reason, it would be meaningless to calculate an average rating across all the jokes for each participant; rather, we wanted to assess the effect of adding laughter to the funniness of each individual joke. We therefore calculated the average rating for each joke across the participants in a group. An independent samples t-test was used to compare the baseline sample ratings of the jokes allocated to the two sets, which indicated that there was no difference in how funny the two sets of jokes were perceived to be ($t(38)=.247$, $p = .806$, Set one: $M = 2.728$, $SD = .551$, $SEM = .123$; Set two: $M=2.688$, $SD = .468$, $SEM=.105$).

We compared the average rating for each joke between the groups using paired sample analyses. A paired samples t-test between the baseline sample and NT group (n=48) showed that jokes with laughter (collapsed across spontaneous and posed laughter) ($M = 3.115$, $SD = .439$) were rated significantly funnier than jokes without laughter ($M=2.708$, $SD = .505$, $t(39) = 6.420$, $p < .001$) (Fig. S2).

We next investigated whether the autism group and NT subgroup differed in the effect that different types of laughter had on the perceived funniness of the joke. To rule out low level attentional effects on performance, we compared the number of failed catch trials in the autism and NT groups using an independent samples t-test, which indicated that there was no difference between the groups ($t(48)= 1.080$, $p = .285$, autism: $M = 1.154$, $SD = 1.120$, $SEM = .220$; NT: $M = .833$, $SD = .963$, $SEM = .197$).

A 2x2 repeated measures analysis of variance (ANOVA) was conducted, including type laughter type (spontaneous vs. posed) and participant group [autism vs. NT (n=48)]as the within-subject factors. There was a significant main effect of type of laughter, $F[1,39] = 22.336$, $p < .001$, $\eta_p^2 = .364$, indicating that jokes with spontaneous laughter ($M = 3.353$, $SD = 0.527$) were rated as more funny than jokes with posed laughter ($M = 3.145$, $SD = 0.539$). There was also a significant main effect of group, $F[1,39] = 19.248$, $p < .001$, $\eta_p^2 = .330$, indicating that the autistic participants ($M = 3.381$, $SD = 0.555$) rated the jokes as significantly funnier than the NT participants ($M = 3.117$, $SD = 0.496$). There was no significant interaction between laughter type and group, $F[1, 39] = .173$, $p = .680$, $\eta_p^2 = .004$, indicating that spontaneous and posed laughter had the same effect on how funny the jokes were perceived to be by the NT and autism participants.

To check that any differences between the autism and NT groups were meaningful, we repeated the analysis with the NT subgroup (n=24). There was a significant main effect of type of laughter, $F[1,39] = 19.018$, $p < .001$, $\eta_p^2 = .328$, indicating that jokes with spontaneous laughter ($M = 3.282$, $SD = 0.548$) were rated as more funny than jokes with posed laughter ($M = 3.073$, $SD = 0.534$). There was also a significant main effect of group, $F[1,39] = 58.075$, $p < .001$, $\eta_p^2 = .598$, indicating that the autistic participants ($M = 3.381$, $SD = 0.555$) rated the jokes as significantly funnier than the NT participants ($M = 2.965$, $SD = 0.492$). There was no significant interaction between laughter type and group, $F[1, 39] = .217$, $p = .644$, $\eta_p^2 = .006$, indicating that spontaneous and posed laughter had the same effect on how funny the jokes were perceived to be by the NT and autism participants.

Supplemental References

S1. Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence (WASI). (San Antonio: Harcourt Assessment).

S2. Weschler, D. (1997). WAIS-III, Wechsler adult intelligence scale, Third Edition (Pearson).

S3. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. J. Autism Dev. Disord, *31*, 5-17.

S4. Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., and Bishop, S.L. (2012). ADOS-2: Autism Diagnostic Observation Schedule, Second Edition. Part 1: Module 4.

S5. Kamp-Becker, I., Ghahreman, M., Heinzel-Gutenbrunner, M., Peters, M., Remschmidt, H., & Becker, K. (2013). Evaluation of the revised algorithm of Autism Diagnostic Observation Schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders. Autism *17*, 87-102.

S6. Lavan, N., Lima, C. F., Harvey, H., Scott, S. K., & McGettigan, C. (2015). I thought that I heard you laughing: Contextual facial expressions modulate the perception of authentic laughter and crying. Cogn. Emotion *29*, 935-944.

S7. McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. Cereb. Cortex *25*, 246-257.

S8. Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer, Version 4.5.16 (Princeton University).

S9. Audacity Team (2018). Audacity: Free Audio Editor and Recorder, Version 2.3.0 retrieved December 20th 2018 from https://audacityteam.org/

S10. Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our Midst: An online behavioral experiment builder. bioRxiv, 438242.