



Assessing Theory of Mind by Humor: The Humor Comprehension and Appreciation Test (ToM-HCAT)

Simge Aykan* and Erhan Nalçacı

Department of Physiology, Ankara University School of Medicine, Ankara, Turkey

OPEN ACCESS

Edited by:

Willibald Ruch,
Universität Zürich, Switzerland

Reviewed by:

Ursula Beermann,
Universität Innsbruck, Austria
Raymond A. Mar,
York University, Canada

*Correspondence:

Simge Aykan
saykan@ankara.edu.tr

Specialty section:

This article was submitted to
Personality and Social Psychology,
a section of the journal
Frontiers in Psychology

Received: 04 April 2018

Accepted: 26 July 2018

Published: 13 August 2018

Citation:

Aykan S and Nalçacı E (2018)
Assessing Theory of Mind by Humor:
The Humor Comprehension
and Appreciation Test (ToM-HCAT).
Front. Psychol. 9:1470.
doi: 10.3389/fpsyg.2018.01470

Theory of Mind (ToM) may be defined as the ability to understand the mental states, such as beliefs, desires, intentions, and emotions, of others. Impairment of ToM ability leads to disorders with pathologies in social skills, such as autism spectrum disorder and schizophrenia. In addition to differences in ToM ability among patient populations, there is variation between neurotypical individuals. Unfortunately, ToM tasks are usually developed for children or patients with cognitive disorders and cannot detect variations in healthy adults. As an alternative tool, humor may be used. Humor plays a role in social communication and requires many different cognitive functions. Humor is believed to represent complex high-order cognitive processes. There are numerous types of humor; the most complex type is considered ToM humor, where an understanding of social/emotional content is necessary. Given the need for a ToM assessment test suitable for healthy adult populations, we developed a test for measuring humor comprehension and appreciation, with and without ToM content (ToM-HCAT). The present ToM-HCAT test is a performance test consisting of cartoons. The test measures perceived funniness, reaction time to perceived funniness decision, and meaning inference. Cartoons were selected after pilot studies involving 44 participants. Subscales were constituted according to expert views and confirmed by confirmatory factor analysis ($N = 135$). Goodness of fit values for the final 35-item test were acceptable to excellent: GFI = 0.97; AGFI = 0.97; NFI = 0.97; RFI = 0.97, and SRMR = 0.067. Both categories were internally consistent ($\alpha_1 = 0.84$, $\alpha_2 = 0.94$). External validity was assessed against autistic traits. One hundred and three participants completed the Autism Spectrum Quotient and were grouped by ± 0.5 standard deviations from the mean as high in autistic traits. The meaning-inference scores of the subscale with the ToM cartoons were significantly lower ($p = 0.034$) for the high autistic traits group, providing evidence of external validity. In conclusion, we developed and validated a test for assessment of ToM by humor comprehension and appreciation. We believe that the present test will be useful for the detection of variations in ToM ability in the healthy adult population.

Keywords: autistic traits, cartoons, humor, reliability, theory of mind, validity

INTRODUCTION

As highly social beings, humans encounter a variety of interactions during their daily lives. Being successful in this environment requires insight into the social and emotional context and understanding of others' intentions and aims, which is enabled by empathy. Empathy can be described as *"Any process that emerges from the fact that observers understand others' states by activating personal, neural and mental representations of that state, including the capacity to be affected by and share the emotional state of another; assess the reasons for the other's state; and identify with the other, adopting his or her perspective"* (de Waal and Preston, 2017, p. 498). In other words, empathy is characterized by the sharing of emotions and consideration of the perspectives of other people. Empathy may be divided into two categories: affective and cognitive (Singer, 2006; Zaki et al., 2012). Affective empathy involves the ability to match others' emotions, while cognitive empathy refers to the ability to imagine how others feel. A type of cognitive empathy is theory of mind (ToM), which may be defined as the ability to understand the mental states of others, such as their beliefs, desires, intentions, and emotions (Wellman and Estes, 1986). In brief, ToM refers to the ability to understand one's own, and others', minds (Baron-Cohen, 2000). The importance of ToM may be illustrated by disorders in which ToM is impaired, such as autism spectrum disorders (ASDs) and schizophrenia (Baron-Cohen, 2000; Bora and Pantelis, 2013; Chung et al., 2014). Regarding autism, one of two main areas of impairment is that of social skills/communication: it is known that ASD patients exhibit the ability to share the emotions of others, but cannot mentalize (Smith, 2006), which is an indicator of impaired cognitive empathy. As underlying causes of disorders, ToM impairments have been studied extensively (Baron-Cohen, 2000; Losh et al., 2012; Chung et al., 2014; Sommer et al., 2018). Regarding schizophrenia, ToM impairments have been shown in unaffected relatives, ultra high-risk individuals, and first-episode patients as evidence of the trait-based nature of the disease (Bora and Pantelis, 2013; Lavoie et al., 2013).

In addition to differences in patient populations, variation in ToM abilities is observed among neurotypical individuals (Baron-Cohen et al., 2001a). However, ToM assessments are usually developed for children or cognitively disabled people (for a review, see Turner and Felisberti, 2017). As a result, the task used may not be sufficiently difficult for healthy adults with strong cognitive and social skills. These tasks have ceiling effects (near 100% accuracy) for healthy control participants (Corcoran et al., 1995; Gallagher et al., 2000; Brüne, 2003; Marjoram et al., 2005), which makes the detection of variations impossible. This limits the investigation of ToM in healthy populations, which is unfortunate as such investigations may shed light on its underlying mechanisms. Although the investigation of a cognitive mechanism in disabled populations provides useful insights, the results may be confounded by the presence of comorbid conditions. In addition, the disabled cognitive mechanism might be compensated by other processes, which will again lead to misinterpretation. Thus, investigation of the mechanism underlying a cognitive process should be

accompanied by research in the healthy population. ToM variations are known to exist in healthy individuals (Baron-Cohen et al., 2001a), examples of whom include healthy first-degree relatives of schizophrenia patients (Janssen et al., 2003; Anselmetti et al., 2009; Bora and Pantelis, 2013) and relatives of individuals with ASD (Baron-Cohen and Hammer, 1997; Losh and Piven, 2007; Gokcen et al., 2009). Further, self-reported and neuroimaging data indicate variance in social cognition in the normal population (Hooker et al., 2010; Wagner et al., 2011; Regenbogen et al., 2015); however, behavioral data are lacking.

Current tasks used to measure ToM vary from social vignettes (e.g., false belief tasks, social animation tasks) to narrative fictional stories and films (e.g., strange stories tasks). There are limited number of ToM ability tests sensitive to variation in healthy population: the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001a), the Faux Pas Test (Stone et al., 1998; Gregory et al., 2002), the Yoni Test (Shamay-Tsoory and Aharon-Peretz, 2007), the DANVA (Nowicki and Duke, 2001), among others. As ToM is a highly complex process with cognitive and affective components that can be implicit and explicit, using a diversity of approaches for assessment is necessary. In addition to current tests, the use of humor represents a potential alternative method for ToM assessment in the healthy population. Humor might be described as anything that people say or do that is perceived as funny and makes others laugh (Martin, 2007). Humor is a way of communicating ideas, strengthening relations, improving group harmony, and expressing aggressiveness in a positive manner. Humor is the most flexible tool for social interaction. Therefore, it is important to express and understand humor to communicate more effectively. Humor is a stimulus encountered often in our daily lives, and the evaluation of humorous material may be considered similar to real-life situations, which makes it an appropriate tool for measuring ToM ability. The simplest form of humor is the pun, which uses visual or semantic resemblance, and the most complex form is ToM humor, which requires ToM abilities (Vrticka et al., 2013).

Humor processing consists of two stages: comprehension (the first stage) and appreciation (the second; Suls, 1972; Wyer and Collins, 1992; Vrticka et al., 2013). The most accepted theory of these is 'incongruity detection and resolution,' which states that humor requires the introduction of the incongruity as a violation of expectations, followed by a resolution associated with enjoyment (Shultz, 1972; Martin, 2007). Humor comprehension requires understanding of the context and detection of incongruity (Ruch, 1992; Coulson et al., 2006; Uekermann et al., 2007). Necessary cognitive processes for incongruity detection may vary from recognition of simple visual resemblance to mentalizing, which requires ToM ability. The second stage, humor appreciation, requires both integration of newly formed meaning in an amusing way and a positive emotional response (Wyer and Collins, 1992; Coulson et al., 2006; Uekermann et al., 2007). Therefore, humor appreciation represents a complex, high order process that involves cognitive, behavioral, physiological, emotional, and social components (Martin, 2007).

In addition to the previously mentioned ToM disability in ASD, another relevant trait is humor impairment. Asperger's

Syndrome (AS; one of the subtypes of ASD) was first defined by Hans Asperger in 1944. Individuals with AS are known to exhibit differences in terms of their perception of humor. This observation is supported by the fact that they have problems in understanding irony or sarcasm (Happé, 1995). Since Asperger's work, researchers have verified humor-related deficits in ASD (Baron-Cohen, 1997; Emerich et al., 2003; Samson and Hegenloh, 2010). Samson and Hegenloh (2010) showed that humor appreciation in ASD individuals depends on the stimulus material. Appreciation was low for ToM cartoons, whereas no difference was observed for visual puns (Samson and Hegenloh, 2010). This result shows that humor appreciation is not reduced when ToM is not necessary.

Humor comprehension and appreciation differences in ASD would be expected to extend to the healthy population with autistic traits. Autistic traits are subthreshold deficits similar to those present in ASD, such as social interaction and communication deficits, as well as restrictive/repetitive behaviors (Constantino and Todd, 2003). The main difference between individuals with ASD and healthy people with autistic traits is the severity of the symptoms. A theory for ASD is that social adaptation and communication skills exhibit a normal distribution among the population, and individuals at the negative end cannot adapt to the social requirements of the population and, thus, constitute the ASD group (Constantino and Todd, 2003; Robinson et al., 2011; Lundström et al., 2012). Accordingly, it is widely known that ASD occurs as a spectrum in the diagnosed population; moreover, this spectrum is also observed among the general population (Baron-Cohen, 1995; Lundström et al., 2012; Ruzich et al., 2015). There is a genetic and biological overlap in the etiology of ASD and autistic traits (Bralten et al., 2018). Therefore, individuals at the end of this spectrum, with a high level of deficits, constitute the ASD group. Consistent with this view, studies examining autistic traits in the healthy population have been increasing in recent years. In addition, studies demonstrating differences in humor styles and appreciation (Eriksson, 2013; Rawlings, 2013) among healthy people with autistic traits similar to those in the ASD population have been reported.

To date, several instruments have been developed to assess the various dimensions of humor. These can be divided into two main groups: questionnaires and performance tests. As questionnaires are not relevant to the aim of this study, they are not discussed here (for a list of questionnaires, refer Ruch, 2007). Approximately 18 performance tests, with different measurement aims covering humor comprehension, appreciation, reasoning, and motivation, are have been constructed (for a list, see Ruch, 2007). None of these tests were developed directly for ToM assessment. However, studies assessing ToM using cartoons and jokes in patient populations and neurotypical individuals, which are not structured and have not been validated, have been reported (Happé et al., 1999; Gallagher et al., 2000; Samson et al., 2008). The stimuli are mostly unstructured cartoons or jokes that are used only in one study, reducing the possibility of replication.

Cartoons may be classified as one particular type of humorous material, i.e., static visual stimuli, and can be described as jokes

in pictorial form (Nilsen and Nilsen, 2000). Cartoons may either consist of both text and pictures, or only pictures. The advantages of cartoons are that they do not depend merely on linguistic abilities, but also enable the depiction of characters' emotions via their facial expressions or body postures. In contrast, in verbal humor, characters' emotions must be described explicitly (Hempelmann and Samson, 2008). Henceforth, we will discuss studies using cartoons for ToM assessment, as our test consisted only of cartoons.

Cartoons with ToM content, which have been used in more than one study, were developed by Gallagher et al. (2000) and Marjoram et al. (2006). The stimuli consisted of cartoons in three categories; ToM, non-ToM, and jumbled pictures. The cartoons were grouped into categories by researchers and applied to 20 people before actual use. Meaning inference was assessed by open-ended questions and scored by a researcher as correct or incorrect. Another unstructured cartoon set was developed by Happé et al. (1999) and Snowden et al. (2003). Similarly to the above-mentioned study, the cartoons were divided into two categories (physical state and ToM) by researchers. Meaning inference for cartoons was assessed by open-ended questions and scored by researchers (Happé et al., 1999). One final example are the cartoons used by Samson et al. (2008). In this study, selected cartoons were pre-examined in several ways. Cartoons were categorized by five people in three categories as puns, those involving ToM, or as semantic, and cartoons with 90% total agreement were put into the related category. Twenty-one participants rated cartoons for funniness, complexity, and originality, with categories balanced regarding these parameters (Samson et al., 2008).

The research to date indicates that cartoon-based ToM assessment may be very useful; however, a structured, reliable, and validated test is currently not available. In addition, humor is a useful tool assessing ToM in healthy adults without a ceiling effect (Adolphs, 2003). Moreover, humor and ToM problems seem to co-occur, as seen in schizophrenia and ASD populations (Bozikas et al., 2007; Samson and Hegenloh, 2010). Hence, a test that measures both humor and ToM would be useful. Measuring both in the same test will provide an opportunity to understand whether these processes are disabled independently or in relation to each other. Finally, there is no structured humor test currently validated for use with a Turkish population.

Based on these demands, in the present study, we aimed to develop a humor test that measures humor comprehension and appreciation using cartoons with and without ToM content. More specifically, we aimed to create a task that: (i) was sensitive to differences in ToM ability in the healthy adult population, without a ceiling effect; (ii) was able to measure humor comprehension and appreciation ability with and without ToM ability; (iii) has adequate psychometric properties, being both reliable and valid; (iv) was objectively scored; and (v) was easy and quick to apply. Cartoons were presented, and time taken to decide whether the cartoon was funny or not (i.e., reaction time), scoring of funniness level (i.e., funniness score), answers for meaning of cartoons (i.e., meaning-inference score) were collected. The test was validated in relation to autistic traits.

METHODS AND RESULTS

Participants

A total of 147 (79 females and 68 males, mean age = 22.56 years, $SD = 4.41$ years), undergraduate or graduate students from different faculties participated. As humor appreciation and comprehension change with aging (Greengross, 2013) we only included younger adults in order to constitute a more homogenous sample. The inclusion criterion was an age ranging between 18 and 35 years, and exclusion criteria were uncorrected visual impairment, a diagnosed neuropsychiatric disorder, and taking neuropsychiatric medication. The study was approved by the Ethical Committee of Ankara University School of Medicine.

Test Development

The study was conducted in a series of four steps; for simplicity, the methods and results for each step are presented together. The first step of test development consisted of the selection of cartoons and piloting. The second step comprised experts grouping the cartoons. In subsequent steps three and four, reliability and validity were analyzed. In the test, three parameters for cartoons were assessed: reaction time (time taken to decide whether the cartoon was funny or not), funniness score (scoring of funniness level), and meaning-inference score (correct answers for meaning of cartoons). Confirmatory factor analysis was performed using AMOS 21.0 (Arbuckle, 2012), and all other analyses were performed using SPSS version 20.0 software (IBM Corp., 2011).

Step 1: Cartoon Selection

As either preference or dislike for sexual cartoons is known to correlate with personality characteristics (Ruch and Hehl, 1998) and was detected in all the factor analytic studies independent of the structural content (Eysenck, 1942; Herzog and Larwin, 1988; Ruch and Hehl, 1998), cartoons with high sexual content were excluded. Cartoons with low sexual, political, and violence content were collected from printed media or the internet. Colored cartoons were converted to black and white to exclude the facilitating effect of color on object recognition (Rossion and Pourtois, 2004), as this might cause a difference between colored and non-colored cartoons' reaction times. Written cartoons with more than 70 characters were excluded as reading speed might have interfered with reaction times.

Participants were given instructions comprising a two-cartoon demo test, before the actual test session. Funniness scores and reaction times were collected using a computer. Funniness evaluation part of the test was presented in a dimly lit, soundproof room using a laptop with a 15.6", 1366 × 768 pixel resolution screen. MATLAB R2013a (MathWorks) with Psychtoolbox 3.0 (Kleiner et al., 2007) was used for presenting the cartoons and to record the reaction times and funniness scores. Participants were instructed as "A cartoon will appear on the screen and click to mouse when you decide if the cartoon is funny or not. Then a second screen will appear with numbers from one to seven, you should rate the funniness using a scale from one not funny

to seven extremely funny." Cartoons were projected on a gray background in randomized order. The time duration between cartoon presentation and mouse click was recorded as reaction time in seconds. The next screen display consisted of numbers from 1 to 7 and the words "Evaluate funniness level." After the funniness level had been chosen, a new cartoon appeared on the screen.

The second part of the test is the meaning-inference test. It is a paper-based test and cartoons are presented in a booklet with one cartoon per page and with the question "Which one of the following represents the meaning of the cartoon most?" followed by four choices (see **Figure 1** for an example).

Step 1.1: Pilot study 1

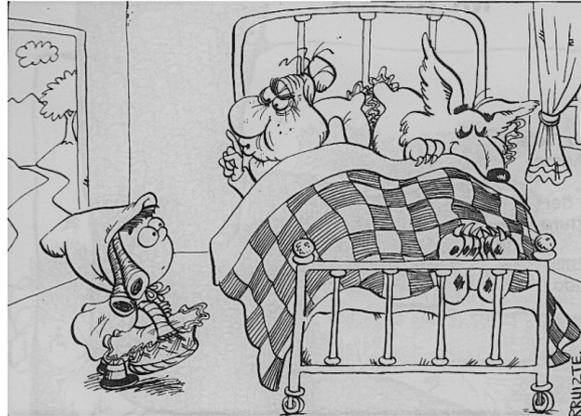
A pilot sample of 12 individuals participated in this step (six females and six males, mean age = 23.75 years, $SD = 3.33$ years). Sixty cartoons were shown to the participants. The mean funniness-score was 3.07 ± 0.84 (range [1.50; 4.92]). Cartoons of funniness scores less than 2.5 ($n = 22$) were eliminated as they were considered to be unfunny for the target population. Mean reaction-time was 7.59 ± 2.30 s (range [3.61; 13.81]). Cartoons with a reaction time of greater than 12 s ($n = 3$) were eliminated due to their longer processing time, which may have indicated that they were more complex than the other cartoons. New cartoons were added to replace those removed. The next step then commenced with 60 cartoons.

Step 1.2: Pilot study 2

Thirty-two people participated in the second pilot study (16 females and 16 males, mean age = 26.63, $SD = 5.11$). Each participant scored the cartoons for funniness. In addition, each participant evaluated 20 cartoons for familiarity and meaning (10 evaluations per cartoon). They then answered two questions: the first question was "Have you seen this cartoon before?" and the second was "Write down the meaning of the cartoon in a single sentence. If you do not think it makes sense at all, you may write down 'meaningless'."

The maximum familiarity was 4/10, and cartoons with a familiarity of 2/10 or more ($n = 20$) were discarded. The study then continued with the remaining 40 cartoons. The mean funniness score and reaction time of each cartoon were calculated. Reaction times with z -scores over ± 3.0 were assigned as outliers as participants might have paused during the test or might have been distracted, and those values were excluded from the analysis. The mean reaction time was 7.03 s ($N = 130-135$, $SD = 1.74$, range [3.72; 10.73]). The mean funniness score was 3.57 ($N = 135$, $SD = 0.46$, range [2.65; 4.60]).

For the first phase of meaning-inference test development, answers to the above mentioned question were collected. In the second phase, four options were created by researchers based on these answers. Four options were designed as follows: one option was the main meaning, two options were secondary meanings, and one option was "meaningless." The place of the main meaning was randomized in the first three choices and the fourth choice was always "meaningless." The main meaning option was considered the correct answer. Every



1. Which one of the following represents the meaning of the cartoon most?

- Grandma does not want the beautifully sleeping wolf to wake up
- Grandma doesn't want the red riding hood to tell anyone that she was in bed with the wolf
- Grandma wants the red riding hood to go away to protect herself from the wolf
- Meaningless

FIGURE 1 | Sample question from meaning-inference test. Option b represents the main meaning. Options a and c are secondary meanings. Option d is "meaningless." Cartoon is reprinted with permission from the Aydın Doğan Foundation. Copyright© 1986, Aydın Doğan Foundation.

correct answer was scored as 1 point. The total number of points was referred to as the "meaning-inference score." The meaning-inference test was presented to participants in two orders in opposite directions to eliminate the possible confounding effect of losing concentration toward the end. The test was applied to a pilot group of 10 participants. For three cartoons, the targeted choice was chosen by fewer than 50% of the participants and, thus, the options have been rearranged.

Step 2: Grouping of Cartoons Depending on ToM Content: The Experts' View

To group cartoons into the two categories as ToM/Non-ToM, nine experts with at least a doctoral degree (social psychology $n = 4$, clinical psychology $n = 2$, developmental psychology $n = 1$, physiology $n = 2$) answered inquired to answer either yes or no to the question: "Do you think that social relations, values, feelings, and thoughts of people need to be understood in order to understand this cartoon?" The number of experts was chosen as an odd number, as this will always result in predominance of either the "yes" or "no" answer. Cartoons were assigned to the ToM category if the majority decision was "yes." Cartoons with a majority of "no" were assigned to the Non-ToM (N-ToM) category. A Mann-Whitney U test indicated that the amount of "yes" votes that were associated with the ToM subscale ($Mdn = 7$) was significantly higher than the amount of "yes" votes that were associated with the N-ToM subscale ($Mdn = 3$), $U = 0$, $p < 0.001$. The ToM group consisted of 27 cartoons, whereas the N-ToM group consisted of 13 cartoons.

Step 3: Reliability

A group of 103 people participated in this part of the study (57 females and 46 males, mean age = 19.68, $SD = 1.85$). The funniness scores of 32 participants, who took part in pilot study 2 were included in the analysis. Therefore, data obtained from 135 participants (73 females and 62 males, mean age = 21.33, $SD = 4.20$ years) were used for reliability analysis.

The reliability of the subscales was assessed by three methods. First, Cronbach's alpha coefficients were calculated as a measure of internal consistency (Cronbach and Meehl, 1955). For Cronbach's alpha, values over 0.70 are accepted as good (Streiner and Norman, 1995; Kline, 2000). The Cronbach's alpha coefficient was 0.84 for the N-ToM group and 0.94 for the ToM group, indicating good internal consistency. Both subscales showed good reliability.

Second, split-half reliability was used as another measure of internal reliability. A basic assumption of split-half reliability is that the two halves of the test should yield similar true scores and error variances (Brown, 1910; Spearman, 1910). In each subgroup, cartoons were divided into two groups (even- and odd-numbered). Spearman-Brown coefficients (r_{sb}) were calculated. The coefficient for the N-ToM group was $r_{sb} = 0.83$, and for the ToM group it was $r_{sb} = 0.95$, indicating good consistency.

The third method involved the calculation of item-total correlations. Descriptive statistics, corrected item-total correlations of all items can be found in **Table 1**. The correlation coefficient is expected to be positive, above 0.30 (Nunnally and Bernstein, 1994). As shown in **Table 1**, item-total correlations

TABLE 1 | Psychometric characteristics of the scale.

Item	Subscale	<i>M</i>	<i>SD</i>	<i>CITC</i>	<i>SRC</i>	<i>SMC</i>
1	N-ToM	3.26	1.89	0.54	0.58	0.34
2	N-ToM	2.91	1.80	0.58	0.64	0.42
3	ToM	2.96	1.72	0.62	0.67	0.44
4	ToM	3.36	1.70	0.62	0.66	0.43
5	ToM	2.72	1.53	0.56	0.60	0.36
6	N-ToM	3.38	1.83	0.59	0.62	0.39
7	ToM	3.99	1.95	0.57	0.59	0.34
8	ToM	3.43	1.95	0.51	0.54	0.29
9	ToM	3.75	1.87	0.68	0.68	0.46
10	ToM	3.69	1.66	0.61	0.61	0.38
11	N-ToM	3.12	1.84	0.48	0.48	0.23
12	ToM	3.34	1.77	0.69	0.71	0.50
13	N-ToM	4.12	1.81	0.45	0.47	0.22
14	ToM	3.50	1.89	0.64	0.66	0.43
15	ToM	3.57	1.82	0.64	0.62	0.39
16	ToM	3.60	1.90	0.62	0.62	0.39
17	ToM	3.86	1.80	0.62	0.62	0.39
18	ToM	3.69	1.94	0.60	0.62	0.39
19	ToM	4.04	1.82	0.70	0.71	0.50
20	N-ToM	3.40	1.88	0.55	0.60	0.36
21	ToM	4.58	1.85	0.52	0.57	0.32
22	ToM	3.56	1.76	0.51	0.54	0.29
23	N-ToM	2.66	1.57	0.48	0.50	0.25
24	ToM	3.81	1.68	0.61	0.62	0.39
25	ToM	3.80	1.90	0.58	0.60	0.35
26	ToM	3.72	1.91	0.59	0.61	0.37
27	N-ToM	3.60	2.00	0.36	0.48	0.23
28	ToM	3.59	1.76	0.48	0.50	0.25
29	ToM	4.57	1.74	0.63	0.64	0.41
30	N-ToM	2.93	1.69	0.58	0.58	0.33
31	N-ToM	4.00	1.77	0.50	0.52	0.28
32	N-ToM	3.28	1.88	0.46	0.51	0.26
33	ToM	4.17	1.97	0.62	0.62	0.39
34	N-ToM	4.04	1.90	0.54	0.60	0.37
35	ToM	3.95	1.89	0.49	0.52	0.27
36	ToM	3.83	1.87	0.64	0.67	0.45
37	ToM	3.42	1.98	0.62	0.64	0.41
38	ToM	2.78	1.68	0.57	0.55	0.31
39	ToM	3.10	1.75	0.57	0.59	0.35
40	N-ToM	3.94	1.81	0.62	0.69	0.48

N = 134; *M*, mean; *SD*, standard deviation; *CITC*, corrected item-total correlation; *SRC*, standardized regression coefficient; *SMC*, squared multiple correlation; *N-ToM*, non-theory of mind; *ToM*, theory of mind.

were above 0.30 and positive, demonstrating the consistency of each item.

Step 4: Validity

Step 4.1: Construct validity

To evaluate the construct validity of the resulting model, confirmatory factor analysis was performed on the funniness scores. Data from 135 participants (the sample that was regarded within the reliability analysis) were used for confirmatory factor analysis.

Bartlett's Sphericity Test and Keiser-Meyer-Olkin (KMO) were calculated as measures of the suitability of data for structure detection. For data to be considered suitable, the Bartlett's test should be significant and the KMO value should be over 0.80 (Bartlett, 1954; Kaiser and Rice, 1974). The data were suitable for factoring as the Bartlett's test was significant ($p < 0.001$) and the KMO value was 0.90.

According to the Mahalanobis distance measure, one participant was detected as a multivariate outlier and, thus, was excluded from the sample (Mahalanobis, 1936). Fit indices were estimated using the unweighted least-squares (ULS) method as

TABLE 2 | Descriptives of test scores and reaction times for main study group.

Score	Subscale											
	N-ToM					ToM						
	Min	Max	M	SD	95% CI		Min	Max	M	SD	95% CI	
					LL	UL					LL	UL
Funniness	10	63	33.25	11.75	30.96	35.55	30	151	91.15	30.57	85.17	97.12
Reaction time ^a	2.11	12.67	7.08	2.11	6.66	7.49	2.49	12.92	7.21	1.74	6.87	7.55
Meaning-inference	2	10	6.67	2.00	6.28	7.06	5	24	17.68	3.53	16.99	18.37

N = 135; N-ToM, non-theory of mind; ToM, theory of mind; Min, minimum; Max, maximum; M, mean; SD, standard deviation; CI, confidence interval; LL, lower limit; UL, upper limit; a, reaction time is in seconds.

TABLE 3 | Descriptives and comparison of ToM-HCAT scores and reaction times for cartoons in subscales.

Score	Subscale											
	N-ToM (<i>n</i> = 10)					ToM (<i>n</i> = 25)						
	Min	Max	M	SD	Median	Min	Max	M	SD	Median	<i>U</i>	<i>p</i>
Funniness	2.66	4.04	3.36	0.49	3.32	2.72	4.58	3.65	0.46	3.69	85.5	0.151
Reaction time ^a	4.37	9.81	6.96	1.97	6.85	3.72	10.13	7.08	1.68	7.49	120.0	0.872
Meaning-inference ^b	47	89	67.3	12.58	69.50	47	90	71.32	13.31	72.00	100.0	0.377

N-ToM, non-theory of mind; ToM, theory of mind; Min, minimum; Max, maximum; M, mean; SD, standard deviation; CI, confidence interval; LL, lower limit; UL, upper limit; a, reaction time is in seconds; b, correct answer percentage of cartoons.

kurtosis (163.23) and critical ratio (16.30) values suggested a non-normal distribution and data were ordinal in structure (de los Angeles Morata-Ramírez and Holgado-Tello, 2013).

The assessment of model fit was based on several indices. The goodness-of-fit (GFI), adjusted goodness-of-fit (AGFI), standardized root mean square residual (SRMR), normed-fit index (NFI), and Bollen's relative fit index (RFI) were used. The absolute fit indices (GFI and AGFI) calculate the proportion of variance that is accounted for by the model covariance. The SRMR shows the difference between the residuals of the sample covariance matrix and the hypothesized covariance model. The NFI shows the fit of the estimated model with the hypothesized model, and RFI considers inconsistency between the two models (Hooper et al., 2008). For GFI, AGFI, NFI, and RFI scores, >0.95 suggests a good fit whereas scores that are >0.80 suggest an acceptable fit (Bentler and Bonett, 1980; Bollen, 1989; Jöreskog and Sörbom, 1996). A SRMR <0.05 suggests a good data-model fit, while <0.08 suggests an acceptable fit (Hu and Bentler, 1999). Fit indices for the initial model are as follows: CFI = 0.97, AGFI = 0.97, NFI = 0.96, RFI = 0.96, SRMR = 0.070. The GFI, AGFI, NFI, and RFI suggested a good fit, and the SRMR suggested an acceptable fit.

Standardized regression coefficients of the variables were expected to be over 0.40, and all items were above that value. Squared multiple correlations should be over 0.30 but can be tolerated toward 0.10 if the other values are acceptable. Standardized regression coefficients and squared multiple correlations for the items are presented in **Table 1**. All regression values were above the expected value of 0.40. Squared multiple correlation values of 10 items were near but below 0.30.

Items with low correlations were excluded from the model one by one, and fit indices were calculated. Items with lower indices were left out of the model. Three cartoons from the N-ToM group and two cartoons from the ToM group were excluded. The final model fit indices are as follows: CFI = 0.97, AGFI = 0.97, NFI = 0.97, RFI = 0.97, SRMR = 0.067. All the indices suggested an acceptable to good fit. The final test consisted of 35 cartoons: 10 cartoons from the N-ToM and 25 cartoons from the ToM.

Descriptives of test scores and reaction times for participants in this study group can be found in **Table 2**.

Descriptive statistics for the ToM-HCAT scores and reaction times, as well as comparisons of the subgroups, are given in **Table 3**. **Table 3** shows that no difference was found for funniness score, reaction time and meaning-inference score between cartoons in ToM and N-ToM subscales.

Step 4.2: External validity

In the main study group, participants (*n* = 103) completed the Turkish version of the Autism Spectrum Quotient (AQ; Kose et al., 2010) for calculation of autistic trait scores in addition to the humor test. The maximum score for the AQ is 50 points; higher scores indicate higher levels of autistic traits (Baron-Cohen et al., 2001b). As the ASD group constitutes the higher end of the distribution for autistic traits (Robinson et al., 2011; Lundström et al., 2012), we adopted a similar approach in the present sample. Participants with AQ scores of +0.5 standard deviations were grouped as the high-autistic traits group (*n* = 37, mean AQ = 24.32, *SD* = 2.21, range [22; 28]). The rest of the population constituted the low-autistic traits group (*n* = 66, mean AQ = 16.58, *SD* = 3.14, range [6; 21]). A chi-square test was

TABLE 4 | Descriptives and comparison of ToM-HCAT scores for high- and low-autistic traits groups.

Score	Low autistic traits (<i>n</i> = 66)					High autistic traits (<i>n</i> = 37)					<i>U</i>	<i>p</i>
	Min	Max	<i>M</i>	<i>SD</i>	Median	Min	Max	<i>M</i>	<i>SD</i>	Median		
Funniness												
N-ToM	10	52	33.67	11.54	34.5	15	63	32.51	12.25	30.0	1105.5	0.427
ToM	30	151	92.79	31.58	97.0	41	140	88.22	28.85	88.0	1106.0	0.429
Meaning-inference												
N-ToM	2	10	6.83	1.94	7.0	3	10	6.38	2.09	7.0	1068.5	0.289
ToM	5	23	18.08	3.68	19.0	11	24	16.97	3.17	17.0	914.5	0.034

N-ToM, non-theory of mind; *ToM*, theory of mind; *Min*, minimum; *Max*, maximum; *M*, mean; *SD*, standard deviation; *CI*, confidence interval; *LL*, lower limit; *UL*, upper limit. Bold text indicates a statistically significant difference with a *p*-value less than 0.05.

TABLE 5 | Descriptives and comparison of reaction times for high- and low-autistic traits groups.

Subcategory	Low autistic traits			High autistic traits			95% CI		<i>t</i>	<i>df</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>LL</i>	<i>UL</i>			
N-ToM	6.78	2.00	66	7.60	2.24	37	-1.70	0.07	-1.85	67.64	0.069
ToM	6.97	1.68	66	7.64	1.80	37	-1.40	0.05	-1.87	70.15	0.066

N-ToM, non-theory of mind; *ToM*, theory of mind; *Min*, minimum; *Max*, maximum; *M*, mean; *SD*, standard deviation; *CI*, confidence interval; *LL*, lower limit; *UL*, upper limit.

performed, and no difference was found for gender between high autistic traits (27:39 [m:f]) and low autistic traits (19:18 [m:f]) groups, $\chi^2(1, N = 103) = 1.05, p = 0.306$. A Mann-Whitney *U* test indicated that age for the low-autistic traits group (*Mdn* = 19) was not significantly different from that for the high autistic traits group (*Mdn* = 20), *U* = 1119.5, *p* = 0.472.

The funniness-score and meaning-inference score on each of the two subscales were calculated as the sum of scores on that category. Scores were compared between groups and the means, 95% confidence intervals and comparisons of ToM-HCAT scores can be found at **Table 4**. **Table 4** shows that the meaning-inference score for the ToM category was lower for the high-autistic traits group (*Mdn* = 17) than for the low-autistic traits group (*Mdn* = 19); *U* = 914.5, *p* = 0.034. There was no difference for N-ToM and ToM funniness scores or N-ToM meaning-inference scores. The Spearman correlation between ToM meaning-inference scores and autistic traits scores was calculated and found to be low and non-significant, $r_s(102) = -0.14, p = 0.163$. To test the robustness of this result the high-end split was further shifted to +1.0 SD. The high-autistic traits group (*n* = 19, mean AQ = 26.16, *SD* = 1.50, range [24; 28]) and the low-autistic traits group (*n* = 84, mean AQ = 17.82, *SD* = 3.68, range [6; 23]) were compared regarding the respective ToM meaning-inference scores. A Mann-Whitney *U* test indicated that the meaning-inference score for the ToM category was lower for the high-autistic traits group (*Mdn* = 17) than for the low-autistic traits group (*Mdn* = 19); *U* = 551.0, *p* = 0.035. Meaning-inference score of the ToM category was compared between females (*Mdn* = 18) and males (*Mdn* = 18) and there was no difference for gender *U* = 1180.0, *p* = 0.383.

Reaction times for both categories were compared between groups. Descriptives and comparison of reaction times for high- and low-autistic traits groups can be found in **Table 5**.

Results showed that the high-autistic traits group had longer reaction times for both subscales; however, no statistical difference existed between the low- and high-autistic traits groups. As reaction times might have been influenced by the number of characters in speech bubbles or by the amount of text in the cartoons, the relationships between text length and reaction time was analyzed. In the N-ToM subscale, 5/10 cartoons had speech bubbles or text. Similarly, in the ToM subscale, there were 25 cartoons, of which 10 featured speech bubbles. A Spearman correlation analysis between character count and reaction time showed a moderate positive correlation, $r_s(34) = 0.39, p = 0.022$. Accordingly, the differences between character counts and reaction times for the N-ToM and ToM subscales were analyzed. A Mann-Whitney *U* test indicated that character count for the ToM subscale (*Mdn* = 0) was not significantly different from that for the N-ToM subscale (*Mdn* = 4.50), *U* = 120.5, *p* = 0.872. Similarly, the reaction time for the ToM subscale (*Mdn* = 7.49) was not significantly different from that for the N-ToM subscale (*Mdn* = 6.85, *U* = 120.0, *p* = 0.872).

DISCUSSION

In this study, we developed and validated a humor test, the ToM-HCAT, to assess humor appreciation and comprehension via the use of cartoons. This test comprises two different subscales: one subscale with ToM content and one subscale without ToM content. This theoretically assumed two-dimensional structure was analyzed by confirmatory factor analysis. The data showed an acceptable-to-good model fit, indicating good construct validity. Reliability measures were good and external validity was evident.

The ToM-HCAT is a performance test consisting of 35 cartoons, and has three outputs: (i) reaction time taken to decide whether the cartoon is funny or not; (ii) funniness score for each cartoon and subscale; and (iii) meaning-inference score for each cartoon and subscale. The reaction time reflects the processing speed of humor appreciation. The funniness score represents humor appreciation, and the meaning-inference score indicates humor comprehension. Within the ToM subscale the meaning-inference score reflects ToM ability by means of humor comprehension.

The test comprises cartoons with or without speech bubbles. In the first subscale, half ($n = 5/10$) of the cartoons have speech bubbles; in the second subscale, 10 out of 25 cartoons had speech bubbles. The distribution of cartoon types in groups is similar. For all cartoons, the text was limited to a maximum of 70 characters to exclude the effect of reading speed on reaction times. Further, this allows for reduced linguistic demands for comprehension. Although there was a moderate correlation between character counts and reaction time, no difference for character counts between subscales was observed, which makes it possible to compare these. All cartoons are black and white to exclude the confounding effect of color, especially on funniness scores and reaction times. The cartoons were chosen randomly from a large pool. The internet, printed cartoon books of Turkish cartoonists, and yearly books of the “Simavi International Cartoon Competition” (1983–1993) were used. Eighty-five cartoons were used in the study, and the final test consists of 35 cartoons; among these, 16 were published by international cartoonists, which also enables adaptation to other cultures.

Funniness decision consists of both humor comprehension and appreciation processes, although for appreciation it is not always necessary to comprehend (e.g., non-sense humor) (Ruch and Hehl, 1998). On the other hand, meaning-inference involves only the humor comprehension process. Accordingly, funniness and meaning-inference scores of ToM-HCAT should be considered as representing linked but different processes. In the meaning-inference test, participants choose the meaning from four options and it might happen that they comprehend the meaning after seeing the choices. In addition, it is possible that they may not have understood the main meaning of the cartoon in the previous funniness test. In our opinion, this does not decrease the importance of either result. This is because a funniness decision can be independent of comprehension; further, individuals may be unable to comprehend even after seeing the choices. Supporting this hypothesis, in the present study none of the participants achieved the maximum score for the meaning-inference test.

In addition to this, there are advantages to using a forced choice test. Current cartoon sets used in studies use subjective evaluations in which the researcher scores participants' open-ended answers (Happé et al., 1999; Gallagher et al., 2000). The fact that the choices were selected by researchers in our study may be questioned; however, the choices were created after collecting explanations from a pilot group. Furthermore, there is no inter-rater reliability problems in the multiple-choice method. Inter-rater reliability refers to how similar the data collected by different

raters are. If raters do not consistently agree in their scoring, then examiner specific factors may contribute unduly to observed score variability (Kline, 2011).

Another output of the test are reaction times for the funniness ratings, which provide the opportunity for evaluating the decision time. Decision time may be affected by cognitive processing speed, serving as a possible indicator of the efficiency of these processes. However, it should be noted that reaction time might have been influenced by numerous factors. For example, the complexity of the cartoons might have influenced the processing time. In the present study, we excluded cartoons for which reaction times were very long as such cartoons might have been overly complex. Another pitfall might have been that participants took a break or were distracted. To prevent this, we excluded reaction times with a very high z -score from the analysis. In our comparison group with autistic traits, reaction times were higher for the high-autistic traits group on both subscales; however, this difference was not significant. Similar results are presented in a cartoon Faux Pas Test. ASD participants took longer than neurotypicals to give their responses independent of cartoon types (Thiébaud et al., 2016). Longer reaction times might be related to the higher detail orientation of individuals with symptoms of autism (Dakin and Frith, 2005; Samson and Hegenloh, 2010). This finding is also an indicator that reaction times may be useful for measuring cognitive processing differences.

In the present study, we showed that individuals with higher autistic traits exhibit poorer humor comprehension if ToM is necessary for understanding the cartoon. It is widely known that the humor response of individuals with ASD differs from the response of neurotypical participants (Van Bourgondien and Mesibov, 1987; Baron-Cohen, 1997; Reddy et al., 2002; Samson and Hegenloh, 2010; Samson et al., 2013). This finding may be interpreted as a result of social communication deficits observed in this disorder (American Psychiatric Association, 2013). Regarding individuals with ASD, the response to humor varies according to the type of humor. Researchers have shown that ASD individuals do not appreciate humor created by socially inappropriate behavior (Reddy et al., 2002), and are unable to readily understand the other person's humorous intention (Baron-Cohen, 1997). High-functioning autistic individuals may make jokes based on lexical or phonological contradictions; however, these tend to be under the age-appropriate level (Van Bourgondien and Mesibov, 1987). In support of our results, a study by Samson and Hegenloh (2010) showed that adults with ASD enjoy visual and semantic pun cartoons at similar levels as neurotypical individuals; however, these individuals exhibit difficulty in understanding ToM cartoons and provide less mentalistic explanations to humor consisting of ToM. In another study, it was shown that adolescents with high-functioning autism or AS performed worse than neurotypical individuals regarding the comprehension of cartoons and jokes (Emerich et al., 2003). We could not show a correlation between meaning-inference scores and AQ; however, this may have arisen from the relatively small sample size. Analysis using a higher number of participants may reveal a significant difference.

As the present population consisted of healthy people with autistic traits without a diagnosis of ASD, this study shows that impairment in ToM and humor extends to the healthy population with autistic traits. In support of this, differences in humor styles and appreciation have been reported among healthy individuals with autistic traits (Eriksson, 2013; Rawlings, 2013). For ToM impairment in healthy individuals, variation was shown by an implicit test: the Reading the Mind in the Eyes test. The test results were negatively correlated with autistic trait scores measured by the AQ (Baron-Cohen et al., 2001a). This result supports the present findings. We found that individuals with higher autistic traits score, as measured by AQ, exhibited poorer comprehension of cartoons with ToM, but not of cartoons without ToM.

To the best of our knowledge, there is no existing test that has the same structure and outputs as our humor test. The most similar test is the 3WD test of humor appreciation (Ruch, 1992): 3WD is another performance test that measures funniness and aversion to cartoons and jokes on a seven-point scale, with 35 items. Three categories of humor are present: non-sense, incongruity-resolution, and sexual. Although the tests are similar in the methods used to measure funniness, there are some differences between the ToM-HCAT and the 3WD test. The most important is that the present test aims to measure ToM processing, and cartoons have accordingly been categorized by their ToM content. To our knowledge, there is no other structured psychometric test that measures ToM ability by humor. A second difference is that we also measured humor comprehension in addition to appreciation.

As our aim was to develop a humor test that measures ToM ability, the finding of lower comprehension scores on ToM subscale cartoons for individuals with high autistic traits supports the validity of our test. Although a difference in funniness scores on the ToM subscale would be expected, we could not find any difference related to autistic traits. In contrast to the present results, a previous study mentioned above reported a significant difference in funniness among AS participants (Samson and Hegenloh, 2010). Although humor comprehension (resolution of incongruity) is considered a prerequisite for humor appreciation (Shultz, 1972; Suls, 1972), it has been suggested that only the detection of incongruity is necessary. This is supported by the appreciation of non-sense or slapstick humor, which does not involve incongruity resolution (Ruch and Hehl, 1998). Therefore, the lack of difference in funniness scores despite the low comprehension scores for the ToM category could be explained by this theory. It is proposed that individuals with high levels of autistic traits find incongruity sufficient for funniness, or that such individuals may perceive a different incongruity and/or resolution. Another difference with the current literature is that we could not show the gender difference in meaning-inference scores for ToM subscale. It is shown that women are superior compared to men in adult ToM tests (Baron-Cohen, 2002). However, in the study by Russell et al. (2007), men showed superior performance compared to women on both physical and mental state cartoons. The results emphasize the hypothesis that the differences in ToM tests could be task specific.

As our starting point was to develop a test to measure variability without a ceiling effect for ToM abilities in the adult healthy population, the findings suggest that our test can detect variability of ToM. Result cannot be attributed to humor ability, because the comprehension scores in the Non-ToM subscale did not show a difference in relation to high or low levels of autistic traits. None of the present participants achieved the perfect score of 25 out of 25 possible points on comprehension for the ToM subscale of the test; further, they scored almost the full range of possible scores of between 5 and 24 points, with a slightly left-skewed distribution. This variation suggests that the ToM-HCAT is sensitive to individual differences in ToM ability. This sensitivity in comparison with other tests may be attributable to the more real-world orientation of cartoons. Cartoons could be regarded as complex social scenarios that require social knowledge, and participants are required to make inferences about their meaning by both explicit mental state reasoning and spontaneous mental state inference. Furthermore, cartoons represent stimuli encountered in daily life.

Limitations

There are several limitations to this study. First, the mean age of participants in the CFA analysis was 21.33 years, with the majority being between 18 and 22 years of age. Second, all participants were undergraduate/graduate students. The use of a more diverse sample is expected to enhance the validity of the current results. In particular, the age range should be wider. Another limitation is the application of the meaning-inference test on paper as a separate test. It may be beneficial to perform this test using a computer to ensure continuity of the entire test. Moreover, reaction times to meaning decisions should be collected, as these may be informative of processing time differences for ToM between individuals with high and low levels of autistic traits. Lastly, studies with larger numbers of participants are required. The present test was not validated using a diagnosed ASD population; this may appear to represent a limitation as this would be a gold standard for ToM disability. However, we validated the ToM-HCAT with autistic traits, which are more subtle than in individuals diagnosed formally with ASD. We showed that the test enables differentiation between these groups, thereby demonstrating its high sensitivity. Moreover, this test was developed to assess variations in ToM ability among the general population. This study, considering the small sample size, should be considered the first step of a scale development process. In future studies, a cross validation phase with a second and larger sample is necessary.

CONCLUSION

In conclusion, a test for assessing ToM involving humor comprehension and appreciation was developed. The item and scale characteristics were good to excellent. The test was externally validated with autistic traits. It has multiple outputs

and is suitable for use in future ToM assessment studies, especially in the healthy population, as it is sensitive to variations in ToM ability among neurotypical individuals. This test is expected to deepen our understanding of differences in ToM ability in the healthy adult population.

AUTHOR CONTRIBUTIONS

SA: design of the study, data collection and analysis, writing and contribution to all parts of the paper. EN: conception and design

REFERENCES

- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4, 165–178. doi: 10.1038/nrn1056
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596
- Anselmetti, S., Bechi, M., Bosia, M., Quarticelli, C., Ermoli, E., Smeraldi, E., et al. (2009). “Theory” of mind impairment in patients affected by schizophrenia and in their parents. *Schizophr. Res.* 115, 278–285. doi: 10.1016/j.schres.2009.09.018
- Arbuckle, J. L. (2012). *Amos (Version 21.0)*. Chicago, IL: IBM SPSS.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S. (1997). Hey! It was just a joke! Understanding propositions and propositional attitudes by normally developing children and children with autism. *ISR. J. Psychiatry Relat. Sci.* 34, 174–178.
- Baron-Cohen, S. (2000). Theory of mind and autism: a review. *Int. Rev. Res. Ment. Retard.* 23, 169–184. doi: 10.1016/S0074-7750(00)80010-5
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn. Sci.* 6, 248–254. doi: 10.1016/S1364-6613(02)01904-6
- Baron-Cohen, S., and Hammer, J. (1997). Parents of children with asperger syndrome: what is the cognitive phenotype? *J. Cogn. Neurosci.* 9, 548–554. doi: 10.1162/jocn.1997.9.4.548
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001a). The “Reading the mind in the eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi: 10.1111/1469-7610.00715
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001b). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471
- Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *J. R. Stat. Soc. Ser. B* 16, 296–298. doi: 10.2307/2984057
- Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606. doi: 10.1037/0033-2909.88.3.588
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons, Inc. doi: 10.1002/9781118619179
- Bora, E., and Pantelis, C. (2013). Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: systematic review and meta-analysis. *Schizophr. Res.* 144, 31–36. doi: 10.1016/j.schres.2012.12.013
- Bozikas, V. P., Kosmidis, M. H., Giannakou, M., Anezoulaki, D., Petrikis, P., Fokas, K., et al. (2007). Humor appreciation deficit in schizophrenia: the relevance of basic neurocognitive functioning. *J. Nerv. Ment. Dis.* 195, 325–331. doi: 10.1097/01.nmd.0000243798.10242.e2
- Bralten, J., van Hulzen, K. J., Martens, M. B., Galesloot, T. E., Arias Vasquez, A., Kiemeneij, L. A., et al. (2018). Autism spectrum disorders and autistic traits share genetics and biology. *Mol. Psychiatry* 23, 1205–1212. doi: 10.1038/mp.2017.98
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Br. J. Psychol.* 3, 296–322. doi: 10.1111/j.2044-8295.1910.tb00207.x

of the study, supervision of all parts of this project, contribution to all parts of the paper. Both authors contributed to the writing of the manuscript, read it critically, and gave consent to its publication.

ACKNOWLEDGMENTS

The authors thank Derya Öztuna for commenting on statistical analysis of the project. The authors also thank the Aydın Doğan Foundation for reprint permission of the cartoon.

- Brüne, M. (2003). Theory of mind and the role of IQ in chronic disorganized schizophrenia. *Schizophr. Res.* 60, 57–64. doi: 10.1016/S0920-9964(02)00162-7
- Chung, Y. S., Barch, D., and Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophr. Bull.* 40, 602–616. doi: 10.1093/schbul/sbt048
- Constantino, J. N., and Todd, R. D. (2003). Autistic traits in the general population. *Arch. Gen. Psychiatry* 60, 524–530. doi: 10.1001/archpsyc.60.5.524
- Corcoran, R., Mercer, G., and Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating “theory of mind” in people with schizophrenia. *Schizophr. Res.* 17, 5–13. doi: 10.1016/0920-9964(95)00024-G
- Coulson, S., Urbach, T. P., and Marta, K. (2006). Looking back: joke comprehension and the space structuring model. *Humor – Int. J. Humor Res.* 19, 229–250. doi: 10.1515/humor.2006.013
- Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957
- Dakin, S., and Frith, U. (2005). Vagaries of visual perception in autism. *Neuron* 48, 497–507. doi: 10.1016/j.neuron.2005.10.018
- de los Angeles Morata-Ramírez, M., and Holgado-Tello, F. (2013). Construct validity of likert scales through confirmatory factor analysis: a simulation study comparing different methods of estimation based on pearson and polychoric correlations. *Int. J. Soc. Sci. Res.* 1, 54–61 doi: 10.11114/ijsss.v1i1.27
- de Waal, F. B. M., and Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nat. Rev. Neurosci.* 18, 498–509. doi: 10.1038/nrn.2017.72
- Emerich, D. M., Creaghead, N. A., Grether, S. M., Murray, D., and Grasha, C. (2003). The comprehension of humorous materials by adolescents with high-functioning autism and Asperger’s syndrome. *J. Autism Dev. Disord.* 33, 253–257. doi: 10.1023/A:1024498232284
- Eriksson, K. (2013). Autism-spectrum traits predict humor styles in the general population. *Humor* 26, 461–475. doi: 10.1515/humor-2013-0030
- Eysenck, H. J. (1942). The appreciation of humour: an experimental and theoretical study. *Br. J. Psychol. Gen. Sect.* 32, 295–309. doi: 10.1111/j.2044-8295.1942.tb01027.x
- Gallagher, H., Happé, F., Brunswick, N., Fletcher, P., Frith, U., and Frith, C. (2000). Reading the mind in cartoons and stories: an fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21. doi: 10.1016/S0028-3932(99)00053-6
- Gokcen, S., Bora, E., Erermis, S., Kesikci, H., and Aydin, C. (2009). Theory of mind and verbal working memory deficits in parents of autistic children. *Psychiatry Res.* 166, 46–53. doi: 10.1016/j.psychres.2007.11.016
- Greengross, G. (2013). Humor and aging – A mini-review. *Gerontology* 59, 448–453. doi: 10.1159/000351005
- Gregory, C., Lough, S., Stone, V., Erzincliglu, S., Martin, L., Baron-Cohen, S., et al. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer’s disease: theoretical and practical implications. *Brain* 125, 752–764. doi: 10.1093/brain/awf079
- Happé, F. G. E. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor Symb. Act.* 10, 275–295. doi: 10.1207/s15327868ms1004_3

- Happé, F., Brownell, H., and Winner, E. (1999). Acquired “theory of mind” impairments following stroke. *Cognition* 70, 211–240. doi: 10.1016/S0010-0277(99)00005-0
- Hempelmann, C. F., and Samson, A. C. (2008). “Cartoons: drawn jokes?,” in *The Primer of Humor Research*, ed. V. Raskin (Berlin: Mouton de Gruyter), 609–640. doi: 10.1515/9783110198492.609
- Herzog, T. R., and Larwin, D. A. (1988). The appreciation of humor in captioned cartoons. *J. Psychol.* 122, 597–607. doi: 10.1080/00223980.1988.9915534
- Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., and D’Esposito, M. (2010). Neural activity during social signal perception correlates with self-reported empathy. *Brain Res.* 1308, 100–113. doi: 10.1016/j.brainres.2009.10.006
- Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural equation modeling: guidelines for determining model fit. *EJBRM.* 6, 53–60.
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118
- IBM Corp. Released (2011). *IBM SPSS Statistics for Windows, Version 20.0.* Armonk, NY: IBM Corp.
- Janssen, I., Krabbendam, L., Jolles, J., and van Os, J. (2003). Alterations in theory of mind in patients with schizophrenia and non-psychotic relatives. *Acta Psychiatr. Scand.* 108, 110–117. doi: 10.1034/j.1600-0447.2003.00092.x
- Jöreskog, K. G., and Sörbom, D. (1996). *LISREL 8: User’s Reference Guide.* Lincolnwood, IL: Scientific Software International.
- Kaiser, H. F., and Rice, J. (1974). Little jiffy, mark IV. *Educ. Psychol. Meas.* 34, 111–117. doi: 10.1177/001316447403400115
- Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What’s new in psychtoolbox-3. *Perception* 36, 1–16
- Kline, P. (2000). *The Handbook of Psychological Testing.* London: Routledge.
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modelling.* New York, NY: The Guilford Press.
- Kose, S., Bora, E., Erermis, S., and Aydın, C. (2010). Psychometric features of Turkish version of autism-spectrum quotient. *Anatol. J. Psychiatry* 11, 253–260.
- Lavoie, M.-A., Plana, I., Bédard Lacroix, J., Godmaire-Duhaime, F., Jackson, P. L., and Achim, A. M. (2013). Social cognition in first-degree relatives of people with schizophrenia: a meta-analysis. *Psychiatry Res.* 209, 129–135. doi: 10.1016/j.psychres.2012.11.037
- Losh, M., and Piven, J. (2007). Social-cognition and the broad autism phenotype: identifying genetically meaningful phenotypes. *J. Child Psychol. Psychiatry* 48, 105–112. doi: 10.1111/j.1469-7610.2006.01594.x
- Losh, M., Martin, G. E., Klusek, J., Hogan-Brown, A. L., and Sideris, J. (2012). Social communication and theory of mind in boys with autism and fragile x syndrome. *Front. Psychol.* 3:266. doi: 10.3389/fpsyg.2012.00266
- Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., et al. (2012). Autism spectrum disorders and autisticlike traits: similar etiology in the extreme end and the normal variation. *Arch. Gen. Psychiatry* 69, 46–52. doi: 10.1001/archgenpsychiatry.2011.144
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Natl. Inst. Sci.* 2, 49–55
- Marjoram, D., Gardner, C., Burns, J., Miller, P., Lawrie, S., and Johnstone, E. (2005). Symptomatology and social inference: a theory of mind study of schizophrenia and psychotic affective disorder. *Cogn. Neuropsychiatry* 10, 347–359. doi: 10.1080/13546800444000092
- Marjoram, D., Job, D. E., Whalley, H. C., Gountouna, V.-E., McIntosh, A. M., Simonotto, E., et al. (2006). A visual joke fMRI investigation into theory of mind and enhanced risk of schizophrenia. *Neuroimage* 31, 1850–1858. doi: 10.1016/j.neuroimage.2006.02.011
- Martin, R. A. (2007). *The Psychology of Humor: An Integrative Approach.* Burlington, MA: Elsevier Academic Press.
- Nilsen, A. P., and Nilsen, D. L. F. (2000). *Encyclopedia of 20th-Century American Humor.* Phoenix, AZ: Oryx Press.
- Nowicki, S., and Duke, M. P. (2001). “Nonverbal receptivity: The Diagnostic Analysis of Nonverbal Accuracy (DANVA),” in *The LEA Series in Personality and Clinical Psychology. Interpersonal Sensitivity: Theory and Measurement*, eds J. A. Hall and F. J. Bernieri (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers), 183–198.
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory.* New York, NY: McGraw-Hill.
- Rawlings, D. (2013). Humor preference and the Autism Quotient in an undergraduate sample. *Humor* 26, 411–421. doi: 10.1515/humor-2013-0028
- Reddy, V., Williams, E., and Vaughan, A. (2002). Sharing humour and laughter in autism and Down’s syndrome. *Br. J. Psychol.* 93, 219–242. doi: 10.1348/000712602162553
- Regenbogen, C., Kellermann, T., Seubert, J., Schneider, D. A., Gur, R. E., Derntl, B., et al. (2015). Neural responses to dynamic multimodal stimuli and pathology-specific impairments of social cognition in schizophrenia and depression. *Br. J. Psychiatry* 206, 198–205. doi: 10.1192/bjp.bp.113.143040
- Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., et al. (2011). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch. Gen. Psychiatry* 68, 1113–1121. doi: 10.1001/archgenpsychiatry.2011.119
- Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart’s object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33, 217–236. doi: 10.1068/p5117
- Ruch, W. (1992). “Assessment of appreciation of humor: Studies with the 3 WD humor test,” in *Advances in Personality Assessment*, Vol. 9, eds C. D. Spielberg and J. N. Butcher (Hillsdale, NJ: Lawrence Erlbaum Associates), 27–75. doi: 10.5167/uzh-77946
- Ruch, W. (2007). “Appendix: humor measurement tools,” in *The Sense of Humor: Explorations of a Personality Characteristic*, ed. W. Ruch (Berlin: De Gruyter), 405–412. doi: 10.1515/9783110804607-019
- Ruch, W., and Hehl, F.-J. (1998). “A two-mode model of humor appreciation: Its relation to aesthetic appreciation and simplicity-complexity of personality,” in *The Sense of Humor: Explorations of a Personality Characteristic*, ed. W. Ruch (Berlin: De Gruyter), 109–142. doi: 10.1515/9783110804607.109
- Russell, T. A., Tchanturia, K., Rahman, Q., and Schmidt, U. (2007). Sex differences in theory of mind: a male advantage on Happé’s “cartoon” task. *Cognit. Emot.* 21, 1554–1564. doi: 10.1080/02699930601117096
- Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., et al. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Mol. Autism* 6:2. doi: 10.1186/2040-2392-6-2
- Samson, A. C., and Hegenloh, M. (2010). Stimulus characteristics affect humor processing in individuals with Asperger syndrome. *J. Autism Dev. Disord.* 40, 438–447. doi: 10.1007/s10803-009-0885-2
- Samson, A. C., Huber, O., and Ruch, W. (2013). Seven decades after Hans Asperger’s observations: a comprehensive study of humor in individuals with Autism Spectrum Disorders. *Humor* 26, 441–460. doi: 10.1515/humor-2013-0026
- Samson, A. C., Zysset, S., and Huber, O. (2008). Cognitive humor processing: different logical mechanisms in nonverbal cartoons—an fMRI study. *Soc. Neurosci.* 3, 125–140. doi: 10.1080/17470910701745858
- Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021
- Shultz, T. R. (1972). The role of incongruity and resolution in children’s appreciation of cartoon humor. *J. Exp. Child Psychol.* 13, 456–477. doi: 10.1016/0022-0965(72)90074-4
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci. Biobehav. Rev.* 30, 855–863. doi: 10.1016/j.neubiorev.2006.06.011
- Smith, A. (2006). Cognitive empathy and emotional empathy in human behavior and evolution. *Psychol. Rec.* 56, 3–21. doi: 10.1080/13682820310001615797
- Snowden, J. S., Gibbons, Z. C., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., et al. (2003). Social cognition in frontotemporal dementia and Huntington’s disease. *Neuropsychologia* 41, 688–701. doi: 10.1016/S0028-3932(02)00221-X
- Sommer, M., Döhnel, K., Jarvers, I., Blaas, L., Singer, M., Nöth, V., et al. (2018). False belief reasoning in adults with and without autistic spectrum disorder: similarities and differences. *Front. Psychol.* 9:183. doi: 10.3389/fpsyg.2018.00183
- Spearman, C. (1910). Correlation calculated from faulty data. *Br. J. Psychol.* 3, 271–295. doi: 10.1111/j.2044-8295.1910.tb00206.x

- Stone, V. E., Baron-Cohen, S., and Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* 10, 640–656. doi: 10.1162/089892998562942
- Streiner, D. L., and Norman, G. R. (1995). *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press.
- Suls, J. M. (1972). “A two-stage model for the appreciation of jokes and cartoons: an information-processing analysis,” in *The Psychology of Humor: Theoretical Perspectives and Empirical Issues*, eds J. H. Goldstein and P. E. McGhee (New York, NY: Academic Press), 81–100. doi: 10.1016/B978-0-12-288950-9.50010-9
- Thiébaud, F. I., White, S. J., Walsh, A., Klargaard, S. K., Wu, H. -C., Rees, G., et al. (2016). Does faux pas detection in adult autism reflect differences in social cognition or decision-making abilities? *J. Autism Dev. Disord.* 46, 103–112. doi: 10.1007/s10803-015-2551-1
- Turner, R., and Felisberti, F. M. (2017). Measuring mindreading: a review of behavioral approaches to testing cognitive and affective mental state attribution in neurologically typical adults. *Front. Psychol.* 8:47. doi: 10.3389/fpsyg.2017.00047
- Uekermann, J., Daum, I., and Channon, S. (2007). Toward a cognitive and social neuroscience of humor processing. *Soc. Cogn.* 25, 553–572. doi: 10.1521/soco.2007.25.4.553
- Van Bourgondien, M. E., and Mesibov, G. B. (1987). Humor in high-functioning autistic adults. *J. Autism Dev. Disord.* 17, 417–424. doi: 10.1007/BF01487070
- Vrticka, P., Black, J. M., and Reiss, A. L. (2013). The neural basis of humour processing. *Nat. Rev. Neurosci.* 14, 860–868. doi: 10.1038/nrn3566
- Wagner, D. D., Kelley, W. M., and Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb. Cortex* 21, 2788–2796. doi: 10.1093/cercor/bhr074
- Wellman, H. M., and Estes, D. (1986). Early understanding of mental entities: a reexamination of childhood realism. *Child Dev.* 57, 910–923. doi: 10.2307/1130367
- Wyer, R. S., and Collins, J. E. (1992). A theory of humor elicitation. *Psychol. Rev.* 99, 663–688. doi: 10.1037/0033-295X.99.4.663
- Zaki, J., Ochsner, K. N., and Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2018 Aykan and Nalçacı. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.