Received: 13 November 2020 | Accepted: 3 May 2021

DOI: 10.1002/aur.2575

RESEARCH ARTICLE

Further developing the Frith–Happé animations: A quicker, more objective, and web-based test of theory of mind for autistic and neurotypical adults

Lucy A. Livingston[1,2] | **Punit Shah**[3] | **Sarah J. White**[4] | **Francesca Happé**[2]

[1]School of Psychology, Cardiff University, Cardiff, UK

[2]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK

[3]Department of Psychology, University of Bath, Bath, UK

[4]Institute of Cognitive Neuroscience, University College London, London, UK

Correspondence
Lucy A. Livingston, School of Psychology, Cardiff University, Cardiff, UK.
Email: livingstonl@cardiff.ac.uk

Punit Shah, Department of Psychology, University of Bath, Bath, UK.
Email: p.shah@bath.ac.uk

Funding information
Cauldron Science; Medical Research Council

Abstract

The Frith–Happé Animations Test, depicting interactions between triangles, is widely used to measure theory of mind (ToM) ability in autism spectrum disorder (ASD). This test began with recording, transcribing, and subjectively scoring participants' verbal descriptions, which consistently found ToM-specific difficulties in ASD. More recently in 2011, White et al. created a more objective version of this ToM test using multiple-choice questions. However, there has been surprisingly little uptake of this test, hence it is currently unclear if White et al.'s findings replicate. Further, the lack of an online version of the test may be hampering its use in large-scale studies and outside of research settings. Addressing these issues, we report the development of a web-based version of the Frith–Happé Animations Test for autistic and neurotypical adults. An online version of the test was developed in a large general population sample (study 1; $N = 285$) and online data were compared with those collected in a lab-based setting (study 2; $N = 339$). The new online test was then administered to adults with a clinical diagnosis of ASD and matched neurotypical controls (study 3; $N = 231$). Results demonstrated that the test could successfully be administered online to autistic adults, who showed ToM difficulties compared to neurotypical adults, replicating White et al.'s findings. Overall, we have developed a quicker, more objective, and web-based version of the Frith–Happé Animations Test that will be useful for social cognition research within and beyond the field of autism, with potential utility for clinical settings.

Lay Summary

Many autistic people find it hard to understand what other people are thinking. There are many tests for this 'mentalising' ability, but they often take a long time to complete and cannot be used outside of research settings. In 2011, scientists used short silent animations of moving shapes to create a fast way to measure mentalising ability. We developed this into an online test to use in research and clinics to measure mentalising ability in autism.

KEYWORDS
autism spectrum disorder, Frith–Happé animations, mentalising, theory of mind, triangles test, web-based research

Researchers interested in using our web-based version of the Frith–Happé Animations Test can either email Sarah White (s.white@ucl.ac.uk) for the video stimuli or Punit Shah (p.shah@bath.ac.uk) who can provide access to the test as programmed on *Gorilla* (Gorilla.sc).

INTRODUCTION

There is considerable evidence that atypical 'theory of mind' (ToM)—the ability to infer other people's mental states (Happé, 2015)—is a cognitive feature of autism spectrum disorder (ASD; e.g., Cantio et al., 2018). A variety of tasks have been developed to measure ToM ability, which have provided evidence for ToM difficulties in autistic children and adults. Initially, ToM was measured in children using the classic false-belief task (e.g., Baron-Cohen et al., 1985), on which autistic children tend to show difficulty in representing a belief that does not correspond to their own view of the world. Following this, many more advanced ToM measures were developed in which participants are required to infer mental states of others from verbal vignettes (e.g., Happé, 1994), pictures of the eye region (e.g., Baron-Cohen, Wheelwright, Hill, et al., 2001) or video-clips of characters interacting (e.g., Dziobek et al., 2006; Murray et al., 2017). There is, however, growing awareness of the limitations of current ToM tasks in autism research, particularly for measuring ToM in adults. First, there are claims of poor validity and suboptimal and subjective scoring, which might be compounded by other cognitive (e.g., verbal, emotional) differences in ASD (e.g., Livingston, Carr, & Shah, 2019; Olderbak et al., 2019). Second, some tasks also produce ceiling effects when administered in neurotypical and autistic adults, and therefore do not capture sufficient variance in task performance (e.g., Happé, 1994). Finally, there are also practical issues with more ecologically valid ToM measures, which are lengthy to administer (e.g., Movie for the Assessment of Social Cognition takes ~40 min; Dziobek et al., 2006; Shah et al., 2017) and require a trained experimenter, limiting their use outside of research settings and in large-scale population-based studies. Together, this has led to suggestions that we should be moving towards abbreviated tasks, involving multiple-choice and automated scoring systems, which can be administered online and/or in clinical settings (Livingston, Carr, & Shah, 2019).

The present study therefore aims to develop a web-based version of a quick, objective test of ToM—called the Frith–Happé Animations Test—adapted by White et al. (2011). The Frith–Happé Animations Test consists of two triangles interacting in one of three ways: drifting or bouncing like objects (Random condition), responding to each other's behaviour (goal-directed; GD), or responding to each other's mental states (ToM). The original version (Abell et al., 2000; see also Castelli et al., 2000)—widely used in autism research (e.g., Livingston et al., 2019)—involves recording, transcribing, and subjectively scoring participants' verbal descriptions of the animations. White et al. (2011) adapted the task to be more objective by using multiple-choice questions, whereby participants select whether each animation depicts 'no interaction' (Random), 'physical interaction' (GD), or 'mental interaction' (ToM). In line with previous findings of atypical ToM in ASD, 16 autistic adults had greater difficulty than 15 neurotypical participants with accurately processing the ToM, but not the Random or GD, animations (White et al., 2011). It was suggested that the objective method was as sensitive as the traditional subjective method in demonstrating well-established ToM difficulties in ASD, making the multiple-choice animations test a more useful research tool.

Despite this progress, we note some areas of White et al.'s (2011) method that could be further developed. First, they did not examine the associations between objective and subjective scores, which we aimed to address to further validate the objective test. Second, there has been little uptake of their test. The objective task is potentially less sensitive to individual differences in ToM, such that researchers may have failed to detect and publish associations between autism and task performance. Therefore, we aimed to replicate White et al.'s (2011) results in larger, more heterogeneous samples. Third, a web-based version of the test could be more efficient. Given the 'replication crisis' in many areas of science, including clinical psychology (Tackett et al., 2017), this would enable collection of larger and more diverse datasets—for example, autistic people who cannot attend labs—and reduce experimenter time. There is an increasing number of web-based platforms that facilitate programming of complex cognitive tasks for online data collection (see Anwyl-Irvine, Massonnié, et al., 2020, for an overview). However, it is currently unclear whether web-based (social) cognitive tasks are feasible and if they perform similarly online to in the lab, as very few studies directly compare online and lab performance (although see Germine et al., 2012). Therefore, we aimed to develop a web-based version of the Frith–Happé Animations Test and, critically, compare performance from online and lab participants. Finally, partly because of the aforementioned limitations, the test is rarely used outside research settings (e.g., clinics), which might become possible through development of a more accessible, online test. Such a task may also be useful for clinicians, where time is limited and a short, objective measure—potentially completed at home or in the clinic—would be advantageous. Overall, there is a need for a quicker, more objective, web-based version of the Frith–Happé Animations Test. Given the rapid increase in online and large-scale research, particularly in the era of COVID-19, this could prove to be a timely and useful task for rapidly measuring ToM in autistic and neurotypical adults outside of the lab. Across three studies, we aimed to develop such a task, and in the process, conduct a fresh empirical test of ToM in ASD.

STUDY 1

Methods

Participants were 285 adults drawn from online sources (aged 18–80 years, M_{age} = 27.98, SD_{age} = 11.99;

174 females) who self-reported levels of autistic traits using the autism-spectrum quotient (AQ; Baron-Cohen, Wheelwright, Skinner, et al., 2001). We adapted White et al.'s (2011) procedure to develop a web-based version using *Gorilla* (Gorilla.sc; Anwyl-Irvine, Massonnié, et al., 2020). There were 12 animations divided into three sets of four animations: 'Random' animations depicted two triangles moving randomly (e.g., drifting, bouncing); GD animations showed two triangles in coordinated physical interaction (e.g., dancing, fighting); 'ToM' animations represented an interaction involving one triangle manipulating the other's mental state (e.g., tricking, persuading).

Informed consent was obtained online from all participants, and all procedures were in line with the local ethics committees, British Psychological Society guidelines, and the 1964 Helsinki declaration and its amendments. Participants were free to withdraw from the study at any time. The study was accessed remotely via a web browser, starting with a definition of each type of animation. Following three practice trials with feedback (one of each animation type), 12 experimental trials were presented in a pseudo-randomised order. Each trial began with the animation auto-playing centrally onscreen (384×288 px). Whilst viewing the animations, participants were required to select if the animation depicted no interaction (Random), physical interaction (GD), or mental interaction (ToM) between the triangles. To prompt intuitive responding, they were instructed to respond as quickly and accurately as possible using on-screen buttons (via mouse press) located below the animation. Only the first response was accepted, with no feedback. The participants viewed the entire animation and then were required to provide a free-text response (via keyboard) to 'what happened in the animation?' before the next trial. Trials were interleaved with a 100 ms fixation cross. The order in which participants completed the AQ and the Animations Test was randomised. All participants completed the test via a web browser on their own computer, rather than a mobile phone or tablet. Recent research has suggested that *Gorilla* is validated for the selection of stimuli via mouse press and that there are minimal influences of browser, devise type or operating system on remotely-collected data that is not time-sensitive (Anwyl-Irvine, Dalmaijer, et al., 2020).

Following White et al. (2011), participants could score a maximum of 12 (4 for each animation type) for objective scores, which were converted into percentage accuracy (Table 1). The free-text descriptions of the animations were reliably scored for the correct inference by three coders (Krippendorff's $\alpha = 0.89$) in line with Castelli et al.'s (2000) 'appropriateness' score. This generated subjective scores between 0 and 8 for each of the animation types, with higher scores indicating greater accuracy (for ToM animations, this means more accurate inference of the triangles' mental states).

RESULTS AND DISCUSSION

A power analysis showed we had at least 80% power to detect 'small-to-medium' associations. All three conditions had acceptable-to-excellent internal consistency (Random: $\alpha = 0.72$, GD: $\alpha = 0.84$, ToM: $\alpha = 0.80$) and data were appropriate for parametric analyses. Objective and subjective scores were significantly correlated (rs = 0.26–0.48; Table 1), thereby providing convergent validity for White et al.'s (2011) objective scoring method.[1] There was also a significant negative association between autistic traits and accuracy on ToM ($r = -0.16$, $p = 0.007$), but neither GD ($r = 0.02$, $p = 0.72$) nor Random ($r = -0.01$, $p = 0.91$) animations. The correlation between autistic traits and ToM animations was small, but significantly greater than the association with GD ($z = -2.79$, $p = 0.005$) and Random ($z = -2.01$, $p = 0.044$) animations.

This pattern of results was in line with previous reports of specific autism-related difficulties in the ToM condition (e.g., Livingston, Carr, & Shah, 2019), providing convergent and divergent validity for our online test. More generally, our findings indicated that, unlike many other ToM tasks, the test is sensitive to individual differences in neurotypical individuals. Therefore, in appropriately large samples, as made possible using the internet, the task may be useful to quantify ToM in the general population.

STUDY 2

Although the results from study 1 suggested that the online version of the task was comparable to previous lab-based studies, there are concerns with the administration of psychological measures online. Whilst some suggest poorer validity (Ramsey et al., 2016), others have found cognitive tasks operate similarly when administered in the lab and online (Germine et al., 2012). To explore this issue, we compared online data from study 1 to lab-based data.

Methods

In addition to study 1's participants, 54 participants (aged 18–41 years, $M_{age} = 24.85$, $SD_{age} = 4.96$; 39 females) formed a convenience sample recruited using a local participant database. These participants undertook the same computerised procedure as study 1, but in a dimly lit, soundproofed laboratory, following experimenter instructions. Lab-based participants were, as expected, younger than study 1 participants, $t(191.34) = 3.19$, $p = 0.002$, $d = 0.34$, given that online

[1]The subjective scores only served to validate the objective measure and are not reported hereafter.

TABLE 1 Means and correlations in study 1

Measure	*M* (SD)	1	2	3	4	5	6
(1) Objective ToM	59.65 (29.05)	-					
(2) Objective GD	63.95 (28.08)	0.39***	-				
(3) Objective Random	83.95 (20.64)	0.19**	0.27***	-			
(4) Subjective ToM	3.29 (2.60)	0.26***	0.33***	0.23***	-		
(5) Subjective GD	4.43 (2.77)	0.21***	0.26***	0.22***	0.76***	-	
(6) Subjective Random	5.70 (2.15)	0.18**	0.24***	0.48***	0.54***	0.60***	-

Note: Mean values are percentage accuracy for the objective measure and accuracy (maximum score = 8) for the subjective measure. Performance across conditions was above chance (i.e., 33%). Correlations are Pearson's *r*.
Abbreviations: GD, goal-directed, ToM, theory of mind.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

data collection allows for more diverse samples (e.g., in age; Anwyl-Irvine, Massonnié, et al., 2020) and the local participant database we recruited from contained university students.

The data were scored following study 1 procedures for objective scoring; that is, participants could score a maximum of 12 (4 for each animation type), which was then converted into percentage accuracy.

Results and discussion

To assess whether online and lab groups differed on the three different conditions, we conducted pre-planned *t* test analyses. Given the group difference in age, we explored differences in task performance between the two samples with and without controlling for age. Lab-based participants were marginally more accurate in the GD condition than online participants ($t[89.38] = -2.83$, $p = 0.006$, $d = 0.38$), but there were small and non-significant differences in the Random condition ($t[337] = -0.26$, $p = 0.80$, $d = 0.04$) and critical ToM condition ($t[337] = 0.83$, $p = 0.41$, $d = 0.12$; Figure 1). This pattern of results held while controlling for participant age (ToM: $F(1, 336) = 0.96$, $p = 0.33$, $\eta p^2 = 0.003$; Random: $F(1, 336) = 0.00$, $p = 0.99$, $\eta p^2 < 0.01$). The small-to-medium difference in the GD condition remained when controlling for age ($F(1, 336) = 5.65$, $p = 0.018$, $\eta p^2 = 0.017$). It is unclear why this was the case but, importantly, there were no group differences on the Random and critical ToM conditions, thus suggesting that the web-based version of the task overall operates similarly to its use in the lab.

STUDY 3

Having developed the web-based Frith–Happé Animations Test in non-clinical samples, we administered the task to autistic adults and matched controls. Although the internet is widely used for questionnaire-based autism research, there is a paucity of knowledge about measuring (social) cognition in this way. Indeed, the current study reports one of the first social cognitive tasks administered to autistic people online (see also Russo-Ponsaran et al., 2019), therefore representing a methodological development of general interest. In line with White et al. (2011), it was predicted that, compared to neurotypical controls, autistic adults would show difficulties in the ToM, but not the GD or Random, condition.

Methods

Seventy-one participants (36 females) aged 18–67 with a formal autism diagnosis were recruited and compared with 160 participants (80 females) aged 18–80 from study 1, selected to ensure that the groups were closely matched in age, sex, and general mental ability (see Table 2 for group characteristics). Neurotypical participants from study 1 were randomly selected until the groups were matched. General mental ability was estimated using the Spot the Word Task (Baddeley et al., 1993), which has

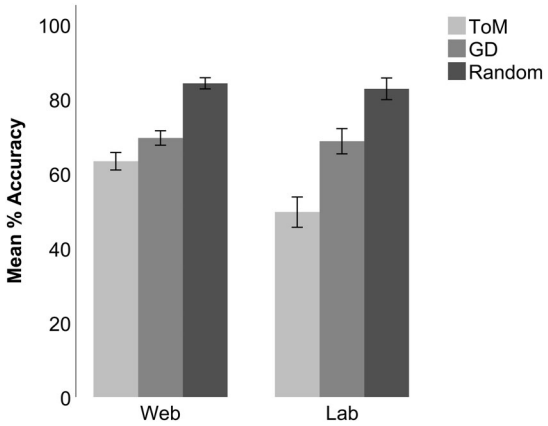

FIGURE 1 Frith–Happé Animations Test—Comparing web and lab performance. ToM, theory of mind; GD, goal-directed. Error bars show ±1 SEM

TABLE 2 ASD and neurotypical group characteristics (study 3)

	ASD ($n = 71$)	Neurotypical ($n = 160$)	
	M (SD)	*M* (SD)	Group comparison
Age (years)	31.48 (11.63)	28.80 (12.05)	$t(229) = -1.58$, $p = 0.12$, $d = 0.23$
General mental ability (% accuracy)	79.91 (22.40)	82.56 (7.54)	$t(77.12) = 0.98$, $p = 0.33$, $d = 0.19$
AQ	**36.41(7.56)**	**17.44(6.93)**	$\mathbf{t(229) = -18.67, p < 0.001, d = 2.66}$
Sex (*n* male, *n* female)	35, 36	80, 80	$\chi^2(1) = 0.01$, $p = 0.92$, $\Phi = 0.01$

Note: General mental ability was estimated using percentage accuracy on the Spot the Word Task (Baddeley et al., 1993). The AQ (autism-spectrum quotient; Baron-Cohen, Wheelwright, Skinner, et al., 2001) measured self-reported autistic traits (maximum score = 50) and has a clinical cut-off of 32+. Effect sizes are reported as Cohen's *d* for *t* tests and Phi Φ for chi squared tests. Significant group differences are shown in bold font.
Abbreviation: ASD, autism spectrum disorder.

previously demonstrated convergent validity with the Wechsler Adult Intelligence Scale (Yuspeh & Vanderploeg, 2000). In this task, participants view 60 pairs of words comprising a real word (e.g., albatross) and non-word (e.g., zando) and are required to identify the real word. Task performance was measured as percentage accuracy. The procedure and objective data scoring were otherwise identical to study 1. Participants accessed the study via *Gorilla* and gave informed consent online and each participant had a percentage accuracy score for each animation type.

Results and discussion

Following White et al.'s (2011) strategy, we compared the groups on ToM, GD, and Random performance (see Figure 2). In line with expectations, on the ToM animations, the ASD group ($M = 49.65$, SD = 33.94) was significantly less accurate than the neurotypical group ($M = 63.28$, SD = 29.80, $t[229] = 3.07$, $p = 0.002$, $d = 0.44$). There were no significant group differences on the GD animations ($t[229] = 0.24$, $p = 0.81$, $d = 0.03$; ASD: $M = 68.66$, SD = 28.57, neurotypical: $M = 69.53$, SD = 24.55) or Random animations ($t[229] = 0.50$, $p = 0.62$, $d = 0.07$; ASD: $M = 82.75$, SD = 24.49, neurotypical: $M = 84.22$, SD = 18.94). Moreover, the group difference in the ToM condition had an effect size similar to that found by Brewer et al. (2017, $d = 0.39$), which, to our knowledge, is the largest *lab-based* ASD-neurotypical comparison on the task. This similarity further reinforces the validity of our online adaptation, indicating its suitability for autistic participants.

We conducted multiple linear regressions to assess the unique contribution of ASD group status to ToM, GD and Random performance, whilst accounting for performance on the other two conditions. These analyses showed that the significant relationship between ASD group and ToM remained even after accounting for GD and Random performance (see Table 3). Further, although our groups were matched on age, sex and general mental ability, because these variables have previously been shown to be associated with ToM ability, we re-conducted the multiple regression analyses with them as additional predictors. We found the same pattern of results. Overall, these regression analyses, not previously undertaken by White et al. (2011), more robustly showed the specificity of ToM difficulties in ASD.

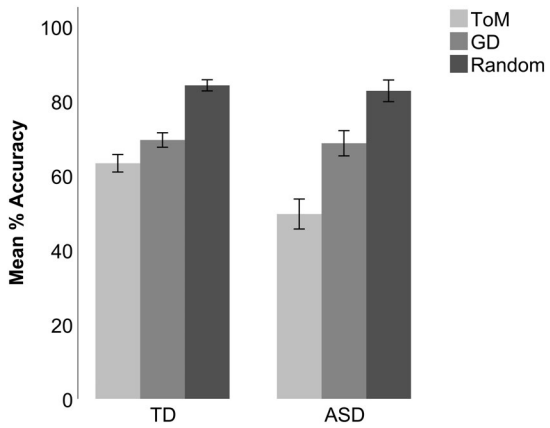

FIGURE 2 Frith–Happé Animations Test—Comparing neurotypical and ASD groups. ASD, autism spectrum disorder; GD, goal-directed; ToM, theory of mind. Error bars show ±1 SEM

GENERAL DISCUSSION

Across three studies, we found that our web-based version of the Frith–Happé Animations Test operates similarly online and in the lab, in both autistic and neurotypical adults. Additionally, we found the expected ToM difficulties in autistic compared to neurotypical adults using online administration. Our findings therefore replicate and extend White et al.'s (2011) finding that the objective version of this popular ToM test is comparable to the traditional version. Enabled by a large sample of the general population, and not directly tested by White et al. (2011), we established that objective and subjective scores collected online were significantly correlated. And importantly, we showed that higher autistic traits were specifically and more strongly linked with online performance on ToM, but not GD or Random, animations.

TABLE 3 Multiple linear regression—Group as a unique predictor of 1) theory of mind (ToM), 2) goal-directed (GD), and 3) random task performance in study 3

	ToM—Overall model fit: $F(3, 227) = 11.31$, $R^2 = 0.13$, $p < 0.001$					
1)	**Predictor**	***B***	**SE *B***	***β***	***t***	***p***
	Group (1 = ASD, 0 = neurotypical)	−13.04	4.24	−0.19	−3.07	0.002
	GD	0.24	0.08	0.20	3.02	0.003
	Random	0.26	0.10	0.17	2.60	0.010
	GD—Overall model fit: $F(3, 227) = 12.52$, $R^2 = 0.14$, $p < 0.001$					
2)	**Predictor**	***B***	**SE *B***	***β***	***t***	***p***
	Group (1 = ASD, 0 = neurotypical)	1.81	3.50	0.03	0.52	0.61
	ToM	0.16	0.05	0.20	3.02	0.003
	Random	0.35	0.08	0.28	4.44	<0.001
	Random—Overall model fit: $F(3, 227) = 11.71$, $R^2 = 0.13$, $p < 0.001$					
3)	**Predictor**	***B***	**SE *B***	***β***	***t***	***p***
	Group (1 = ASD, 0 = neurotypical)	0.24	2.83	<0.01	0.09	0.93
	ToM	0.11	0.04	0.17	2.60	0.01
	GD	0.23	0.05	0.28	4.44	<0.001

Note: All VIF values were <10, suggesting multicollinearity was not a concern. The residuals were normally distributed and there was no evidence of homoscedasticity. Durbin-Watson values were all ~2, suggesting errors were independent. This pattern of results held when including age, sex, and general mental ability as additional predictors in all three regression models but are not reported as the autistic and non-autistic groups were already matched on these variables.
Abbreviations: *β*, standardised regression coefficient; ASD, autism spectrum disorder; *B*, unstandardised regression coefficient; GD, goal-directed; ToM, theory of mind.

Further, in line with White et al. (2011), we found significant differences between autistic and neurotypical people, but only in the ToM condition. This adds weight to ToM theories of autism and indicates that our online test is sufficiently sensitive to detect atypical ToM in intellectually able autistic, as well as neurotypical, adults. This is important as many other ToM tasks appear to be solved by autistic people using compensatory strategies (Livingston et al., 2021; Livingston & Happé, 2017) and/or yield ceiling effects for neurotypical adults. Therefore, we suggest that this test has important utility for future research on ToM in autistic and neurotypical adults. For example, moving forward, the test can now be used to investigate important relationships between ToM and other psychological (e.g., mental health) and social-cognitive (e.g., empathy) constructs within and beyond the field of autism, in large samples and with remote data collection.

Our findings also support suggestions that (social) cognitive research is possible using the internet. Like Germine et al. (2012), we found that participants performed similarly online to in the lab. This finding mitigates concerns about online cognitive research, such as task performance being affected by distractions and/or the lack of experimenter oversight. More generally, this test is one of the first social cognitive tasks to be successfully and specifically developed for online use in both typical and autistic adults. This development—of an objective, quick, online test of ToM—will enable its future inclusion in large scale studies that have traditionally been unable to incorporate lengthy social cognitive tasks (e.g., longitudinal studies, including behavioural genetic studies). This will enable statistically powerful investigations of ToM and its inter-relationships, including genetic correlations, with other psychological constructs and phenotypes across the lifespan. More broadly, this study highlights the opportunities of moving more cognitive autism research online to include 'hard-to-reach' autistic individuals, who may be unable to attend labs, thereby making research more representative of the population (although we note the need to develop ToM tests accessible to autistic people with language/intellectual impairment). Finally, the test can also now be adopted in clinical research to begin assessing its clinical utility (see also, Livingston, Carr, & Shah, 2019). For example, this objective test, which can feasibly be administered prior to a time-limited clinical session, may be useful for clinicians to aid understanding of autistic people's ToM abilities and thereby inform and tailor support, although this needs robust investigation.

Our findings should be considered in light of some limitations. First, we note that across the studies, although we did not formally test this, mean values suggest that neurotypical participants performed better on the Random compared to ToM animations. This differs from White et al. (2011) who found equivalent performance for neurotypical participants on these two animation types. However, we also note that Brewer et al. (2017) found a similar pattern of results to ours when using the task in a much larger lab-based study. Therefore, it is possible that the ToM animations are genuinely more difficult to solve than the Random

animations, which is understandable given the increased complexity of the ToM animations, but that this was not revealed in White et al.'s (2011) small sample. Overall, the critical distinction to make may be between the GD and ToM conditions, given they are more closely matched on complexity and kinematics. Second, although our autistic and neurotypical participants were matched on general mental ability using an online task, future research should aim to replicate our findings using more in-depth measures of IQ. Finally, whilst the current research validated the web-based version of the task in autistic and neurotypical participants, we were not able to test whether performance on the task predicts performance on other ToM tasks, self-report measures of ToM (e.g., Clutterbuck et al., 2021), or everyday social abilities/differences. Future research should aim to investigate, for example, if autistic participants' ToM performance indexes performance on a range of other ToM tasks, as well as autistic behaviour (e.g., using the Autism Diagnostic Observation Schedule; Lord et al., 2000) and social difficulties in the real world.

To conclude, we have developed a new web-based version of the Frith–Happé Animations Test using White et al.'s (2011) multiple-choice version. It performs just as well online as in the lab and shows sensitivity to the measurement of individuals differences in ToM in both autistic and neurotypical adults. There is promise for this web-based test, which offers a fast and straightforward measure of ToM in autistic and neurotypical adults, to be used in future research and clinical work.

ACKNOWLEDGMENTS

L. A. L. was supported by the Medical Research Council, United Kingdom. P. S. was supported by a grant from Cauldron Science. The authors thank Bethany Carr, Holly Bainbridge, Caitlin Foster, and Joseph Riddell, for assistance with data collection.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Lucy A. Livingston https://orcid.org/0000-0002-8597-6525
Punit Shah https://orcid.org/0000-0001-5497-4765
Sarah J. White https://orcid.org/0000-0001-6946-9155

REFERENCES

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Anwyl-Irvine, A. L., Massonnié, J. K., Flitton, A., Kirkham, N., & Evershed, J. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavioral Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01501-5

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The spot-the-word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55–65. https://doi.org/10.1111/j.2044-8260.1993.tb01027.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. https://doi.org/10.1023/a:1005653411471

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring theory of mind in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*(7), 1927–1941. https://doi.org/10.1007/s10803-017-3080-x

Cantio, C., White, S., Madsen, G. F., Bilenberg, N., & Jepsen, J. R. M. (2018). Do cognitive deficits persist into adolescence in autism? *Autism Research*, *11*(9), 1229–1238. https://doi.org/10.1002/aur.1976

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*(3), 314–325. https://doi.org/10.1006/nimg.2000.0612.

Clutterbuck, R. A., Callan, M. J., Taylor, E. C., Livingston, L. A., & Shah, P. (2021). Development and validation of the four-item mentalising index. *Psychological Assessment*, *33*(7), 629–636. https://doi.org/10.1037/pas0001004

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636. https://doi.org/10.1007/s10803-006-0107-0

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Happé, F. (2015). Autism as a neurodevelopmental disorder of mind-reading. *Journal of the British Academy*, *3*, 197–209. https://doi.org/10.5871/jba/003.197.

Livingston, L. A., Carr, B., & Shah, P. (2019). Recent advances and new directions in measuring theory of mind in autistic adults. *Journal of Autism and Developmental Disorders*, *49*(4), 1738–1744. https://doi.org/10.1007/s10803-018-3823-3

Livingston, L. A., Colvert, E., Social Relationships Study Team, Bolton, P., & Happé, F. (2019). Good social skills despite poor theory of mind: Exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *60*(1), 102–110. https://doi.org/10.1111/jcpp.12886

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, *80*, 729–742. https://doi.org/10.1016/j.neubiorev.2017.06.005

Livingston, L. A., Kumarendran, S., & Shah, P. (2021). Definition: Compensation. *Cortex*, *134*, 365. https://doi.org/10.1016/j.cortex.2020.11.002

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., Hammond, N., Murphy, D., & Happé, F. (2017). A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Research*, *10*(5), 1120–1132. https://doi.org/10.1002/aur.1744.

Olderbak, S., Geiger, M., & Wilhelm, O. (2019). A call for revamping socio-emotional ability research in autism. *Behavioral and Brain Sciences*, *42*, e104. https://doi.org/10.1017/S0140525X1800239X

Ramsey, S. R., Thompson, K. L., McKenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior*, *58*, 354–360. https://doi.org/10.1016/j.chb.2015.12.049

Russo-Ponsaran, N. M., Lerner, M. D., McKown, C., Weber, R. J., Karls, A., Kang, E., & Sommer, S. L. (2019). Web-based assessment of social–emotional skills in school-aged youth with autism spectrum disorder. *Autism Research*, *12*(8), 1260–1271. https://doi.org/10.1002/aur.2123

Shah, P., Catmur, C., & Bird, G. (2017). From heart to mind: Linking interoception, emotion, and theory of mind. *Cortex*, *93*, 220–223. https://doi.org/10.1016/j.cortex.2017.02.010

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F., & Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, *12*(5), 742–756. https://doi.org/10.1177/1745691617690042

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of theory of mind for adults with autism. *Autism Research*, *4*(2), 149–154. https://doi.org/10.1002/aur.174

Yuspeh, R. L., & Vanderploeg, R. D. (2000). Spot-the-word: A measure for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology*, *15*(4), 319–326. https://doi.org/10.1016/S0887-6177(99)00020-7.

How to cite this article: Livingston, L. A., Shah, P., White, S. J., & Happé, F. (2021). Further developing the Frith–Happé animations: A quicker, more objective, and web-based test of theory of mind for autistic and neurotypical adults. *Autism Research*, 1–8. https://doi.org/10.1002/aur.2575