

Confabulation in children with autism

Abbreviated form of title: Confabulation in autism

David Spitzer¹, Sarah J White¹, Will Mandy², and Paul W Burgess¹

¹Institute of Cognitive Neuroscience

²UCL Research Department of Clinical, Education & Health Psychology

UCL (University College London), London, WC1E 6BT, UK.

¹Corresponding author: Paul W. Burgess, UCL Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N3AZ. p.burgess@ucl.ac.uk

Publication status

Article title: Confabulation in Children with Autism

Journal title: CORTEX (in press)

Accepted: 5th October 2016

Abstract

Some children with high-functioning autistic spectrum conditions (ASC) have been noted clinically to produce accounts and responses akin to confabulations in neurological patients. Neurological confabulation is typically associated with abnormalities of the frontal lobes and related structures, and some forms have been linked to poor performance on source monitoring and executive function tasks. ASC has also been linked to atypical development of the frontal lobes, and impaired performance on source monitoring and executive tasks. But confabulation in autism has not to our knowledge previously been examined experimentally. So we investigated whether patterns of confabulation in autism might share similarities with neurologically-based confabulation. Tests of confabulation elicitation, source monitoring (reality monitoring, plus temporal and task context memory) and executive function were administered to four adolescents with ASC who had previously been noted to confabulate spontaneously in everyday life. Scores were compared to a typically developing (TD) and an ASC control group. One confabulating participant was significantly impaired at reality monitoring, and one was significantly worse at a task context test, relative to both the ASC and TD controls. Three of the confabulators showed impairment on measures of executive function (Brixton test; Cognitive Estimates test; Hayling Test B errors) relative to both control groups. Three were significantly poorer than the TD controls on two others (Hayling A and B times), but the ASC control group was also significantly slower at this test than the TD controls. Compared to TD controls, two of the four confabulating participants produced an abnormal number of confabulations during a confabulation elicitation questionnaire, where the ASC controls and TD controls did not differ from each other. These results raise the possibility that in at least some cases, confabulation in autism may be less related to social factors than it is to impaired source memory or poor executive function.

Keywords: autism; memory; confabulation; source memory; reality monitoring; executive function

Introduction

Over the last 30 years, the cognitive impairments underlying autism have been increasingly investigated and a number of prominent theories have been suggested, the best known of which is the frequent presence of a mentalizing impairment (Frith, 2003). These cognitive impairments attempt to explain the core behavioural symptoms of the disorder. However, there are also a number of behavioural features that are not considered to be core to the autism spectrum. These occur with some consistency, but only in a minority of individuals with the diagnosis. One such feature is confabulation, which is defined as the tendency to relate experiences that have not occurred, as if they had. Interestingly, some of the first clinical accounts of people with autism noted the occurrence of confabulation. For instance, Hans Asperger described confabulatory behaviour in one of the four children he originally studied:

He was said to be an inveterate 'liar'. He did not lie in order to get out of something that he had done - this was certainly not his problem, as he always told the truth very brazenly - but he told long, fantastic stories, his confabulations becoming ever more strange and incoherent. He liked to tell fantastic stories, in which he always appeared as the hero. He would tell his mother how he was praised by the teacher in front of the class, and other similar tales. (Asperger [1944] translated in Frith, 1991, p.51.)

The tendency for some people with autism to construct such confabulations is recognised within the widely-used assessment tool the Autism Diagnostic Observation Schedule (ADOS-G, Lord et al, 2000) which asks raters to code whether the interviewee “describes an event that seems unlikely to have been real”. Despite this, no research to date has been conducted into confabulatory behaviour in autism. So in this novel research study, we extended a recent trend for applying methods from the field of adult cognitive neuroscience to examine developmental confabulation in children with autism, in particular focusing on source monitoring and executive function.

The nature of confabulation

Confabulation has been defined as “a falsification of memory occurring in clear consciousness” (Berlyne, 1972). People who confabulate are typically agnosognosic: they are unaware of any memory deficit, or at least grossly underestimate its severity (McGlynn and Schachter, 1989). They have no intent to deceive the listener; as such, confabulation has been described by Moscovitch (1989) as “honest lying”. Burgess and Shallice (1996b), describe the principal features of confabulation as follows:

- a. Confabulations are typically (but not necessarily) coherent accounts that concern the person describing them, most apparent when autobiographical recollection is required.
- b. The account is false within its context, and may also be false in detail.
- c. The content is most often drawn from the patient's own experiences, including thoughts that they have had. However some aspects may be derived from semantic memory.
- d. Confabulations are a product of impaired memory processes, not a result of conscious compensatory procedures.
- e. People often act on the basis of their confabulations.
- f. Confabulators are unaware that they are confabulating.

But although these characteristics are common to all confabulations (at least, those identified so far), there are still a very wide variety of different forms, and discussions about them have a substantial history (e.g. Berlyne, 1972; Damasio et al, 1985; Schnider et al, 1996a Kopelman, 1987, 1999; Burgess and McNeil, 1999; Moulin, 2013; Schnider et al, 2013). In general, it is assumed that at least some confabulatory phenomena relate to unusual behaviours seen in experimental tasks that tap "executive" or "metacognitive" aspects of memory. Examples are intrusion errors or false positives in recall of word lists; being willing to answer direct questions to which one cannot possibly know the answer; overconfidence in judging their memory performance, and reality monitoring deficits (e.g. Moulin, 2013; Stolzenberg and Pezdek, 2013; Nahum et al, 2012). There is plenty of evidence that confabulators tend to do (although do not invariably do) poorly on a range of memory tests. Many also perform poorly on tests of executive abilities associated with frontal lobe function. However, the precise relation between different forms of confabulation and performance on experimental or clinical tests is far from clear (e.g. Nahum et al, 2012; Shingakia et al, 2016). It is probably fairest to summarise our current level of understanding by saying that while *most* confabulators show some kind of disturbance on clinical and lab-based memory tests, and/or also often also on tests of executive function, it is not possible to predict well from knowing the test scores what form of confabulation will be displayed, nor perhaps even if confabulation will be shown. Nevertheless, although there is a distinct lack of specificity at present in our understanding of the relation between test performance and symptom, the consistent presence of at least some sort of decrement on tasks relating to memory control, metacognitive and executive functions, albeit unpredictable, means that it is common in experimental investigations of confabulation in neurological patients to include such

measures. Accordingly, we do so here. In particular, we consider explanations that relate to three of the principal putative explanations of confabulation. It is important to note that these accounts are not competitors, nor are they exclusive. Rather, they seek to explain different aspects of the highly varied presentations seen in confabulation, and more than one can be relevant in any one case.

The first, and one of the oldest, accounts is the *dysexecutive hypothesis* of confabulation, which holds that deficits in executive function are a predisposing feature of confabulation (Berlyne, 1972; Stuss et al, 1978; Baddeley and Wilson, 1988). Confabulation is thus viewed as one symptom of the broader set of dysexecutive symptoms (see Burgess, 1998 for overview), and is associated with a wide variety of pathologies, but especially those that affect the frontal lobes (e.g. Turner, Cipolotti, Yousry and Shallice, 2008b). Indeed, Gilboa and Moscovitch (2002), reviewing thirty nine studies, found that 81% of confabulators had damage to prefrontal cortex, most commonly the orbitofrontal and ventromedial aspects of the frontal lobe. According to the dysexecutive hypothesis, a variety of deficits could cause confabulation, e.g. failure to inhibit incorrect responses, inability to monitor behaviour, misuse of environmental cues, a tendency to impulsivity, and lack of concern about correct performance (Stuss et al, 1978; Shapiro et al, 1981). However, since the correlation between confabulation and other dysexecutive symptoms is modest (e.g. Burgess et al 1998 found correlations of between .30 to .54 in a group of 92 patients), and not universally found (see e.g. Schnider and colleagues, 1996b; 2000; Dalla Barba et al, 1990), it is possible that only a small proportion of people who confabulate do so because of executive-type difficulties. It is also possible that there are executive processes that are quite specific to certain forms of memory retrieval, and so they are manifesting only in the domain of confabulation (see Cunningham et al, 1997). Indeed, these memory-specific control processes may be the same as those highlighted by the other two accounts described below, thus highlighting the relations between them. The principal difference in this account is perhaps therefore the idea that contributory factors to the presentation can be measured by tasks that do not prima facie have a strong memory load, but instead measure aspects of thinking such as inhibition, concept formation, reasoning and so forth. Accordingly, we measure these aspects here.

Confabulation has also been looked at from the point of view of *reality monitoring*. Reality monitoring is the ability to distinguish between actual/perceived and imagined events (Johnson, O'Connor and Cantor, 1997; Johnson and Raye, 1998), and is a subform of the broader category of *source memory* (which is knowing where or from whom some information came from rather than what the information was). According to this accounts, confabulation is explained in relation to mistaking imagined events for real ones. For example, Johnson et al (1997) studied a patient who reported that items were presented as words when she had imagined them as pictures; Ward and

Parkin (2000), reported a patient who tended to say he had performed actions he had only imagined; see also Turner et al (2008a). In healthy subjects, real and imagined events can be distinguished by the different sources attributed to each (real events will have, for example, a perceptual base to act as their source, which imagined events will lack); perhaps this ability is deficient in confabulation. This account is appealingly straightforward, but as with all other accounts of confabulation, is unlikely to explain every presentation. For instance, Johnson et al (1997) reported a confabulating frontal patient who did not perform any more poorly at source monitoring than non-confabulating frontal controls, but performed more poorly in other tasks. They concluded that the confabulation was not due to impaired source monitoring alone, but in conjunction with a vivid imagination and an inability to retrieve autobiographical memories systematically. Another limitation is that since this theory concentrates on memory retrieval, it does not explain why particularly bizarre confabulations can be produced in some cases (Dab et al, 1999; Dalla Barba, 1993). Another further difficulty is the prediction that imagined events will be mistaken for real ones; Dalla Barba et al (1997) found in one case that real events were just as likely to be mistaken for imagined ones. Nevertheless, this account is highly appealing in explaining certain forms of confabulation, and has considerable experimental support. So, here we also look at how well experimental tasks measuring this ability might relate to confabulation in our autism participants.

A third prominent account for confabulation is the “temporal context hypothesis”. Temporal context is the ability to recall the correct order of events, and their temporal relations to each other. It is often viewed as a particular type of source memory (e.g. Turner et al, 2008c). This account follows from the observation that the content of confabulations seem to be temporally-loose combinations of real memories, which can often be traced to events in the patient’s past. It has therefore been proposed that confabulation stems from an inability to recognise the temporal context of memories (Schnider et al, 1996b; Schnider, 2003; and Dalla Barba and La Corte, 2015, for a related account). The likelihood that this explanation can explain all forms of confabulation is probably not greater than for the dysexecutive or reality monitoring accounts. There is certainly evidence in favour: for instance, Schnider and Ptak (1999) asked participants to observe stimuli and then, following a break of thirty minutes, they were asked to identify recurrences of stimuli within that session but to forget that they had seen them half an hour previously. Spontaneous confabulators (i.e. those who produce confabulations in everyday life, without necessarily being asked a direct question) identified images they had seen thirty minutes previously as having been seen only moments before, thus supporting the hypothesis. Also concordant with this explanation are observations that confabulators can perform poorly at discriminating when stimuli were presented (Schnider et al, 1996a; Johnson et al, 1997). However it has also been noted that non-confabulators (e.g. amnesics and those with dorsofrontal lesions with no associated memory

problems) can also fail such tasks (e.g. Shimamura et al, 1990; Johnson et al, 1997), and it is hard to account for the bizarre or fixed content of some confabulations through this explanation. A variant of the temporal context hypothesis is based upon patients who confabulate not only to questions relating to the past but also to those relating to the present and future (Dalla Barba et al, 1997). According to this explanation, confabulation stems from impaired awareness of the flow of time in memory (“personal temporality”); this may explain why confabulations often occur in the episodic rather than semantic domain (Dalla Barba et al, 1997). Some confabulators, at least, seem to confuse ongoing reality with the past, so we measure this tendency here also.

So far, we have considered confabulation from the point of view of neurological patients, who frequently show disturbances of memory. But is there evidence for memory alterations in autism? There is indeed some evidence of altered memory function in autism from experimental studies. Individuals with autism tend to perform better than controls in rote-memory tasks in which attention to detail is advantageous, such as the ability to remember random word-lists, but show weaknesses when remembering information in context, such as memory of meaningful sentences (Hermelin and O’Connor, 1967). They also tend to show an enhanced ability to recognise false memories possibly due to a lack of reliance on context, for example when presented with related words (e.g. sharp, pin, sewing) they are better able to recognise similar words not previously presented as novel words (e.g. needle; Beversdorf et al, 2000 but see Bowler et al., 2000 for a lack of replication). Children with autism spectrum conditions (ASC) have been shown to have reduced autobiographical memory (Bruck, London, Landa and Goodman, 2007); proposed explanations include delayed self-recognition impairing their ability to form autobiographical memories (e.g. Howe and Courage, 1997), or delayed social and language development (Nelson and Fivush, 2004). However, children with autism are more likely to make errors of omission than errors of commission (i.e. insertion of events or details) compared to controls (Bruck et al., 2007). In agreement with notions of changes in use of context in autism, autistic children have been shown to be impaired at source monitoring. In a paradigm involving discrimination between cards placed by the subject and cards placed by the experimenter, autistic children made more errors than controls (Russell and Jarrold, 1999). Similarly, Hala, Rasmussen and Henderson (2005) found that autistic children performed more poorly than controls at three types of source monitoring: internal, external and reality monitoring (although see Farrant, Blades and Boucher, 1998 for a lack of replication on a reality-monitoring task).

Similarly, it is well-documented that autistic individuals display perturbations in performance on the same tests of executive function that can be performed poorly by frontal-lobe damaged patients (see e.g. Frith, 2003; Towgood, Meuwese, Gilbert, Turner and Burgess, 2009; White,

Burgess and Hill, 2009). Functional imaging studies have also reported atypical frontal activity in people with a diagnosis of autism when performing executive tasks (e.g. Luna et al, 2002; Gilbert, Bird, Brindley, Frith and Burgess, 2008), and abnormal structure and function of some areas of PFC are often quoted as a potential source of the clinical presentation (e.g. Castelli, Frith, Happé, and Frith, 2002).

On these grounds it might be proposed that there may be some similarities between confabulation of acquired neurological origin, and confabulation of developmental origin. Accordingly in the current study, we identified a small number of high-functioning autistic children who presented clinically with unusual verbal accounts and responses which, if they were in neurological patients, would have been classified as confabulation. We investigated whether these developmental confabulations might share characteristics with neurological confabulations by administering tests of source monitoring and executive function that have elicited poor performance in neurological confabulators, as well as attempting to elicit confabulations. The participants identified as producing confabulations were treated as case studies, and were compared with two control groups: those with and without autism, allowing us to differentiate between impairments general to autism and those specific to individuals with autism who confabulate. This is the first time, to our knowledge, that confabulation in autism has been studied experimentally, and in such detail.

Method

Participants

Ethics approval for the study was received from the Joint UCL/UCLH Committee on the Ethics of Human Research and from the Great Ormond Street Hospital (GOSH) Research Ethics Committee and written consent was obtained from the parents of all participants, or the participants themselves where appropriate, prior to inclusion in the study. Four high-functioning adolescents with a clinical diagnosis on the autism spectrum and a history of confabulation were recruited to take part in the study. Three of them were recruited from GOSH, which offers a specialist national service which is designed for the assessment of complex cases and the provision of second opinions when there is uncertainty at local level as to whether a child has ASD. The service is specifically for the assessment of high-functioning ASD, and so only reflects referrals from children with fluent language, and with current or recent engagement with mainstream education. Participants were identified via discussion with the clinical team, who were asked to identify recent clients who met criteria for ASD and showed confabulation. One further confabulating participant was found when their parents contacted one of the research team (PWB) directly after an internet search of confabulation researchers. We also tested two control groups who had no known history of

confabulation: those with and without ASC, hereon referred to as ASC controls (n=17) and TD controls (n=18). The ASC controls were recruited from participant volunteer databases at UCL and Great Ormond Street Hospital (GOSH). The UCL ones were recruited from mainstream schools (including those with specialist autism units) in the Surrey or Sutton area of the UK, who had a diagnosis of an ASD. They were self-selecting in that every child with a diagnosis at every school was contacted, and the ones we tested were those who were willing to participate. The TD controls were children of friends and neighbours of the first author, and for whom the parents had confirmed that there had never been any concern their children had an autism spectrum, or other developmental, condition. All participants were between the ages of 10 and 17 years and the two control groups were comparable for age ($t(33) = 1.66, p = .107$) and gender ($X^2(1)=.957, p=.328$). Intellectual performance was not measured for the TD controls but is likely to be comparable: while the ASC controls' IQs were non-significantly lower than average (VIQ: $t(13)=1.66, p=.122$; PIQ: $t(13)=1.92, p=.077$), the TD controls were also non-significantly younger, hence the groups were likely of comparable cognitive ability. All 4 confabulating participants were comparable to the ASC controls (within 2 SDs of this group's mean) in terms of age, IQ and symptomatology (see Table 1).

TABLE 1 HERE

Participants identified as producing confabulations.

Participant 1 (P1). P1 was 14 when he took part in this project. He was diagnosed with autism at the age of 3 years and had a history of language delay. The majority of his education had been in mainstream environments with support, most likely due to his average intelligence (VIQ 96, PIQ 88; measured at 10 years with WISC-III-UK, Wechsler, 1992). At age 11, a parental interview (3DI; Skuse et al., 2004) provided evidence of significant difficulties in all three areas of the autistic triad (see Table 1). Aged 13, he scored above the clinical threshold on the ADOS-G (Lord et al., 2000).

During test sessions at age 10, P1 gave incorrect answers seemingly on purpose and pretended to fall off his chair to avoid completing a task. His speech contained unusual phrases, such as "I've only got less than a minute to diffuse any bomb" when he was running out of time on a task, and delayed echolalia mostly referring to films and cartoon characters. During the ADOS at age 13, P1 was noted to be overconfident, overfamiliar and to make various bizarre statements. For example, when asked how he feels inside when he's very happy, he replied "like I'm living the American dream, in heaven, no health and safety restrictions, no embarrassing video cameras on me", and later "I'm afraid of the paparazzi because they want to take photos of you at every angle", both spoken in a casual tone and scattered amongst otherwise fairly typical conversation. During this assessment, it also became noticeable that P1 had a tendency to make statements that were assumed to be confabulatory. Examples are given below.

Example 1. "I've got a crazy Dad who keeps saying I eat too many Christmas crackers"

[S.W. responds in surprise]

"Sometimes I do. He says, er, he keeps on trying to chuck roman candles into my neighbour's front, er, back garden coz I still remember a couple of years earlier when, um, in 2004 we had this huge party and we knocked their roman candle and there's like flares chasing us round the garden and I just ran inside and a second later there's a firework in the house and I'm just like "no, no", and I just like ran upstairs"

[P1 laughs. S.W. asks if it was his Dad who set the firework off.]

"Yeah, um, and what happened was, is that ten seconds later we launched the firework and the fuse blew it up and everyone was coated in soil."

[P1 laughs again]

Example 2. When talking about his favourite day whilst on holiday:

"When we went to that five star hotel I found a white piano and I'm like a grade 4 pianist right now, so I just started playing and the second I start playing I see thirty police officers coming in and I feel like I'm going to get arrested or play a wrong note."

[P1 laughs. S.W. asks what the police officers were doing]

"They were just like having a lunch break, and I never realised, and of course I counted them all up and there were like thirty, forty, fifty, more than I thought! And I decided what on earth is going on, I don't know what's going to happen."

Example 3. When talking about a girlfriend:

"I asked her out and she fainted literally."

[S.W. responds "really?"]

"Yeah and I had to resuscitate her."

Participant 2 (P2). P2 was 11 when she took part in this project. She was diagnosed with autism at the age of 6 years, labelled as 'high-functioning' and educated in mainstream settings. At age 10, she obtained a standardised score of 73 on the Raven's Standard Progressive Matrices, placing her in the Borderline range of intelligence. Physically, she was noted to be over-developed in terms of both height and head circumference (more than 2 and 3SDs respectively above the mean for her age and gender) and was reported to have entered puberty early. Unfortunately, no measures of autism symptom severity were obtained.

P2 first came to our attention at age 10 when her family reported that she frequently related events that had not occurred, resulting in serious consequences on more than one occasion. For instance, she had recounted a detailed story involving inappropriate contact between her and another person while in a car. But when the reported events were investigated, it was proved that they could never have happened. Her mother and older brother were interviewed informally, both

together and separately. Both reported similar patterns of confabulation, which was apparently motiveless. It was unclear whether P2 believed her own false accounts, but her family had the impression that she did, given that these had often led to negative consequences for herself, and they did not regard them as attention-seeking.

Participant 3 (P3). P3 was 17 when she took part in this project. She was given a diagnosis of Asperger's syndrome with attention deficit hyperactivity disorder (ADHD) at the age of 7 years. Her education took place within mainstream settings due to her average to high-average intelligence (VIQ 99, PIQ 117; measured at 8 years with WISC-III-UK, Wechsler, 1992). At age 8, a parental interview (3DI; Skuse et al., 2004) provided evidence of significant difficulties in the first two areas of the autistic triad (see Table 1).

In conversation, P3 came across as being older than she was, accompanied by an insistence that she was much older than she really was. This had begun around the age of 6 and, when seen at 8 years of age, she claimed to be 18 years old. At school she would tell teachers that she was in the classroom as an agent of the headmaster and, during test sessions, she claimed she went to school because she was at teacher training college and was gaining experience. At this age, she also claimed to have had three children by a man who worked at a local restaurant, would play at breastfeeding her babies, and would talk about how painful childbirth was. She had changed the date of birth on documents to make herself ten years older.

During assessment at age 8, P1 was attentive and cooperative although somewhat precocious. She engaged in adult-like conversation and would compliment the work of adults as if she was a peer, not a child. She had a considerable range of gestures and non-verbal communication, although these were somewhat odd and staged. She once described the other children at school as "my small buddies" whilst signing quotation marks in the air with her fingers. Her parents interpreted her obsession with being an adult as an autistic focused-interest/preoccupation.

Participant 4 (P4). P4 was 11 when she took part in this project and had only just received a diagnosis of ASC, including a history of language delay. There was a family history of developmental delay and obsessive compulsive disorder and P4 herself had a history of behavioural difficulties, memory problems and concentration difficulties. Her education had taken place in mainstream settings although her intelligence test scores placed her in the average to low-average range and she was reported to struggle at school (Verbal Comprehension=77; Perceptual Reasoning=98; Working Memory=80; Processing Speed=70; WISC-IV (Wechsler, 2004). At age 11, a parental interview (3DI; Skuse et al., 2004) provided evidence of significant difficulties in all three areas of the autistic triad. At the same time-point, the ADOS-G (Lord et al., 2000) also categorised her as having an ASC (Table 1).

During test sessions at age 11, P4 was highly anxious and avoidant, and would not leave her mother. At times during these sessions she would convey, mostly via her facial expression, a sense that her assessor was behaving somewhat foolishly, for example by adopting a scornful countenance when asked a question. This tended to happen when P4 was asked to attempt activities that were challenging for her. P4 demonstrated no obvious confabulation during her assessment, but her mother spontaneously reported a history of incidents when P4 had told stories that caused considerable alarm, and which on investigation turned out to be false. For example, P4 had previously reported to teachers that she had been physically assaulted by her mother, but investigation had revealed that this was not the case. In fact, it would be more accurate to say the reverse was true, as P4 was reported to have regularly hit her mother and to have threatened a sibling with a weapon on at least one occasion.

Procedure

Confabulation elicitation questionnaire.

This consisted of 55 questions presented in a random order; 5 questions probed each of eleven domains. These covered semantic and episodic information in various time frames, relating to the participant and his/her mother (see Table 2). There were two categories in which the appropriate response was “I don’t know”; participants were instructed that they should not guess if they did not know the answer, but should say “I don’t know”.

TABLE 2 HERE

After testing, the answers were verified with a parent, and the number of confabulations calculated. Occasionally it was difficult to decide whether to attribute an incorrect answer to confabulation, to simple guessing or to genuinely incorrect knowledge. In these circumstances the parent’s impression was used as a guide. Small errors were allowed (e.g. +/- 2 years for “How old is your mum?”)

Source monitoring tasks.

Three computer-based source monitoring tasks, using easily-recognisable images of familiar objects, were administered in the following order: (1) Temporal-context – first study phase (2) Reality monitoring – study and test phase (3) Temporal-context – second study and test phase (4) Task-context – study and test phase. Reaction times and error rates were recorded.

Temporal context task. In the study phase, two sets of 12 images were presented, separated by an interval of approximately 10 minutes. Participants were required to identify whether each item would normally be found inside a house within 10 seconds of the stimulus onset. The test phase immediately followed the second set of study phase images and 36 images were presented in

a pseudorandomised order: the 24 images previously presented plus 12 novel images. Participants were required to identify whether they had seen the picture during the first or second study stage or whether it was new within 10 seconds of the stimulus onset.

Reality monitoring task. This task was based on Aleman, Bocker, Hijman, de Haan and Kahn (2003). The study phase consisted of 32 trials in which a target word was presented, half accompanied by a picture of that word (perceive), the others by a question mark (imagine). Participants were encouraged to picture the target word in the 'imagine' condition. These two types of stimuli were alternated sequentially and participants were asked to name the colour of the item they saw or imagined within 10 seconds of this stimulus onset. The test phase was then immediately administered, consisting of 48 stimuli in a pseudorandomised order: the 32 words previously presented plus 16 novel words (see Figure 1). Participants were required to identify whether each word had previously been perceived, imagined or was new within 10 seconds of the stimulus onset. The number of errors made to previously perceived or imagined stimuli was recorded, as well as the number of imagined stimuli that were identified as having been perceived, and the number of novel items identified as having been previously perceived.

Task context task. The study phase consisted of 24 trials in which participants viewed an image. Prior to this image, they were required to identify within 10 seconds either whether the item could make a noise (12 trials) or whether the item was larger than the testing laptop (12 trials), presented in a pseudo-randomised order. The test phase immediately followed the study phase, consisting of 36 trials: the 24 images seen previously plus 12 new images. Participants were required to identify which question had been asked about each image, or whether the image was new within 10 seconds.

Episodic memory tests. In order to determine whether confabulation occurs in the context of a more widespread memory problem the ASC participants, visual recognition memory was tested using the Doors element and visual recall was tested using the Shapes element of the Doors and People test (Baddeley et al, 1994). The Doors and Shapes task scores were combined to give a total visual-memory score out of 60.

In order to gain a control measure for general word learning effects in the source monitoring tasks, we used a word list learning procedure. First, List A (containing 12 words) was read to the participants. Immediately afterwards, the participants then recalled as many words as they could (the order was not important). Then they were read a second list (List B; also containing 12 words), and recalled as many words as they could. Next, participants were read a list of 40 words (the 12 words from List A, the 12 words from List B, plus 16 new "distractors"). After each word, the participant were asked to say whether they thought the word had been on a list (they did not have to specify which), by answering yes or no. If they said they recognised one of the distractor words as having been on a list, this was classified that as an intrusion, or false-recognition. This procedure

yielded two measures: (a) recall: the number of words correctly recalled in stages (1) and (2), maximum of 24; (b) recognition: the number of words in Stage (3) correctly recognised as appearing on Lists A & B: max. of 24. These scores were combined to give a score out of 48, (listed as “recall and recognition” in Table 3); (c) the number of intrusions/false recognitions (marked out of 16). This is referred to as “New -> Old” in Table 3.

Tests of executive function.

Three measures of executive function were employed: The Hayling Test, the Brixton Test, and the Cognitive Estimates Test.

The Hayling Sentence Completion Test (Burgess and Shallice, 1997) was previously modified for use with children by White et al. (2009); here we administered the task according to the manual (Burgess & Shallice, 1997) but with the child stimuli (White et al., 2009). In Part A, each sentence must be completed with an appropriate word, and in Part B with an unrelated word. Parts A and B therefore both require verbal generation but Part B also requires the inhibition of prepotent verbal responses. Accuracy and response time were recorded for parts A and B separately.

The Brixton Spatial Anticipation Test (Burgess and Shallice, 1997) is a measure of rule attainment and switching that has been shown to be sensitive to frontal lobe dysfunction, and to consequent bizarre responding (Burgess and Shallice, 1996a). In this test, the participant is presented with an array of ten dots, arranged in a 2x5 matrix, with each position numbered (1 is top left, 10 is bottom right). One of the dots is coloured blue and he/she is required to predict the position of the blue dot on the next trial and adjust his/her responses accordingly when the rule changes. An example of a rule change is where the blue dots are following a sequence over several trials of increasing their position by +1, and then switches to a spatial sequence alternating between the top and bottom rows of the matrix instead. The total number of errors in predictions is recorded.

The Cognitive Estimates Test (Shallice and Evans, 1978) was modified for use with children by Happé, Booth, Charlton and Hughes (2006) and this version was used here. The participant was asked a series of 10 questions that were likely to be novel (e.g. how tall are most lamp posts?) to which they were required to make a reasonable estimate. This is thought to tap response selection and monitoring, as well as inhibition. Responses were scored on a four-point scale (0-3) according to data from Happé et al., depending on age group (8-11, 11-13 and adult); 10 year olds were scored according to 8-11 data, those aged 11-15 according to 11-13 data, and those aged 16-17 according to adult data. A total score out of 30 was calculated, as well as a score out of 10 for very extreme responses.

Statistical Analysis

The performance of each confabulation participant was compared to the performances of both the TD and ASC control groups using the procedures for comparing single case scores with groups described by Crawford and colleagues (Crawford and Garthwaite, 2002; Crawford and Howell, 1998). The ASC and TD control groups were also compared to assess whether poor performance in the confabulation participants was specific to those participants or typical of individuals with autism more generally. As the majority of data was not normally distributed in both control groups, non-parametric tests were used. For comparisons of individuals within a control group with others within that same group, $Z = -2$ is used as the cut-off for noteworthiness.

RESULTS

Confabulation Elicitation Questionnaire

Compared to both the TD controls and the ASC controls, P2 produced an abnormal number of confabulations on the Confabulation Elicitation Questionnaire (CEQ) (for comparison with TD controls, two-tailed probability = 0.00023, effect size = 4.769 95% CI 3.060 to 6.466; for comparison with ASC controls, two-tailed probability = 0.011, effect Size = 2.947 (95% CI 1.821 to 4.056)). For P1, the comparison with ASC controls just failed to reach significance ($p = 0.084$, 2-tailed), but the contrast between P1 and the TDs achieved significance (effect size 3.23, 95% CI 2.02-4.43, $p < .01$). P1 produced 5 confabulations on the CEQ, which is the same as the worst TD participant (out of $N = 18$), and the two worst ASC controls ($N = 17$). P1 also produced 2 additional confabulations spontaneously, which are outlined below, but not included in the statistics for the Confabulation Elicitation Questionnaire (CEQ). There was no significant difference between the TD and ASC control groups on this measure ($U = 133.50$, $p = .525$; see Table 3).

TABLE 3 HERE

Here are examples of the participants' confabulations during the confabulation elicitation procedure, presented verbatim (experimenter's dialogue is presented in italics) (these confabulations as additional to those presented for the questionnaire in Table 3):

Confabulating ASC participant 1

[laughter]

Why are you laughing?

I'm just laughing because my friend bought a computer from the Apple store, and it fell into the Thames. He sat on a bench and put it on the rail, he tripped and his head caught the computer, and it fell in.

What is this friend called?

I can't remember his name now.

Are you still friendly with him?

Yes.

Where was this?

The River Thames, near the Millennium Bridge.

What happened then?

He went crazy with me. I lent him five quid to get it. He went crazy, he chased me all around to Trafalgar Square and St James's Park. But I ran into the tube station and got on the tube and he chased me but then the door closed. He videoed it on his phone.

What?

He videoed himself chasing me all the way through the centre of London.

When did this happen?

Don't know, I can't remember. About two years ago.

How can you get a computer for five quid?

I don't know, he was just crazy.

What about the money?

He still wants me to give it to him. I lent him the five quid and he paid me back just five days later.

Note: This appears to be spontaneous confabulation, which began in response to an enquiry as to why the subject was laughing. The subject's mother was not aware of these events, and had not previously heard her son make reference to them. She reported that a friend had recently bought accessories from an Apple store, but the other details sounded unlikely to her.

Similarly, the following spontaneous account began with the subject laughing to himself. According to the subject's mother, it was based on a television programme and did not involve friends of the subject:

[laughter]

Why are you laughing?

Nothing, just a video I saw on YouTube.

What was it?

An old friend, his chums made a slingshot and slung him all the way over the fence, they were meant to sling him into a swimming pool and they smashed him on to the neighbour's Ferrari.

Who were these people?

I don't know, one of them was a friend of mine.

Additionally, the following confabulations were not spontaneous, but were provoked by the confabulation elicitation procedure:

Who was the first person you ever saw on television?

Jonathan Ross. My mum still hates me for that, she sent me upstairs because of the swearing and I watched it on the computer.

How old were you

About eight.

Did you not have a TV before that?

Before that, our electricity was always faulty, then it was fixed and we got a TV.

This answer contains confabulatory details. According to the participant's mother, the family had only had a computer for a year, and their electricity was not faulty.

Do you have any cousins? What are their names and ages?

No.

The participant's mother reported that he had many cousins, but that he confuses them with brothers and sisters. However, this was followed by another incorrect response:

Does your mum have any other children? What are their names and ages?

No.

The participant's mother reported that she did in fact have another child.

Participant 2

When did you last travel by train?

One month ago, to London.

The subject's mother reported that she had not travelled by train for eight months.

Do you have any cousins? What are their names and ages?

No.

According to her mother, the subject has many cousins, whom she has met.

What age did you start school?

One.

This is an unusual response (which according to the mother cannot be explained by the child having gone to playgroup). Only one other participant (a TD control) gave a similar answer.

Does your mum have any other children? What are their names and ages?

No

The participant has siblings.

Who is Bob the Builder?

A builder.

This was the only response (out of all participants and controls) which did not identify Bob the Builder as a cartoon character (it is a very popular children's TV programme in the UK).

Participant 4

What was the name of your first teacher at school?

Linda.

This was incorrect: Linda had worked in the office.

Source Monitoring Tasks (all statistical comparisons are 2-tailed)

Reality monitoring task

P1 made significantly more errors when attempting to correctly identify previously perceived and imagined stimuli than either the TD or ASC control groups (reality monitoring errors; P1 vs TD effect size 3.73, 95% CI 2.35-5.10, $p < .005$; P1 vs ASC controls, two-tailed probability = 0.034, effect size 2.39, 95% CI 1.43 to 3.323). This was mostly due to identifying imagined stimuli as having been perceived (I->P errors, P1 vs TD effect size 3.32, 95% CI 2.11-4.51, $p < .005$; P1 vs ASC

controls two-tailed probability = 0.051, effect size (Z-CC) 2.18 (95% CI 1.28 to 3.05). P1 was also much more likely than the TD controls to identify novel stimuli as having been previously perceived or imagined (i.e. false recognitions; P1 vs TD effect size 7.3 95% CI 4.84-9.81, $p < .00001$). However the difference between P1 and the ASC controls on this measure did not reach significance ($p = 0.38$, 2-tailed). The two control groups did not differ significantly on any of these measures.

Temporal context task

The confabulating participants performed at similar levels to TD controls, except for making false-recognitions where P1 and P4 performed poorly (two-tailed probability for both = 0.0096, effect size 3.0 (95% CI 1.891 to 4.093)). These participants' scores were not significantly different from the ASC controls, and there was no significant difference between the control groups for any of the temporal context measures.

Task context task

P3 performed at similar levels to TD controls (results for P2 were not recorded due to technical failure). But P4 made a significantly larger number of context errors than the TD controls (two-tailed probability = 0.018, effect size 2.682 (95% CI 1.668 to 3.679)) or the ASC controls (two-tailed probability = 0.049; effect size 2.182 (95% CI 1.285 to 3.059)). P1 made a significant number of errors, relative to the TD group, where he identified stimuli he had already seen as being new (two-tailed probability = 0.0085, effect size 3.059 (95% CI 1.932 to 4.169)). Mann-Whitney tests showed no significant difference between the two control groups for any result.

Control tests for source monitoring tasks

Performance of the confabulators on the visual memory and list-learning was generally comparable to TD controls. However P4 performed poorly at recall and recognition in the word-list learning test (comparison with TD controls: two-tailed probability = 0.013, effect size -2.84 (95% CI -3.88 to -1.78)). However, this participant was not significantly poorer than the ASC controls, who were as a group significantly poorer than the TD controls (Mann-Whitney $U = 64.50$, $p = .003$). Indeed, four ASC controls performed as badly as or worse than P4. There was no significant difference between the control groups for Doors and Shapes, and although the ASC group were slightly weaker than the TD group in terms of intrusions/false-recognitions on the list learning task, this did not reach significance.

Test of Executive Function

As Table 3 shows, the difference between ASC and TD controls did not reach significance for Hayling Sentence Completion Test A (verbal initiation speed; Mann-Whitney U , $p = 0.062$; range of ASC controls 5.5-21.1 secs; range of TDs 5.5-15.4). However, they did perform significantly more poorly than TD controls on Hayling B (verbal response suppression; range of ASC controls 7.1 to 104.8 secs; range of TD controls 5.8-10.2) ($U = 68.00$, $p = .008$) and B-A ($U = 79.00$, $p = .024$). There was no significant difference between the two control groups on the remaining tests of executive

function tests however: Hayling B errors ($U=113$, $p = 0.195$), Brixton ($U = 146.50$, $p = .832$), Cognitive Estimates score ($U = 131.00$, $p = .483$), Cognitive Estimates very extreme estimates ($U = 138.50$, $p = .636$).

Within the confabulating ASC participants, P1's performance on the Hayling B (response suppression) condition was significantly different from the TD controls (Hayling B two-tailed probability = 0.00006; effect size 5.57 (95% CI 3.597 to 7.535); B-A two-tailed probability = 0.0018, effect size 3.839 (95% CI 2.432 to 5.231)). This was also the case for P2 (Hayling B two-tailed probability = 0.0001, effect size 5.18 (95% CI 3.388 to 6.958), Hayling B-A two-tailed probability = 0.0031, effect size 3.58 (95% CI 2.256 to 4.890)) and P4 (Hayling B time two-tailed probability = 0.00009, effect size 5.25 (95% CI 3.437 to 7.052); Hayling B-A two-tailed probability = 0.0004, effect size 4.55 (95% CI 2.911 to 6.173)). P2 also produced significantly more errors on the response suppression part of the Hayling Test than either the TD or ASC controls (Part B errors: comparison with TDs, two-tailed probability = 0.007, effect size 3.15 (95% CI 1.994 to 4.285); comparison with ASC controls $p = 0.008$, effect size 3.10 (96% CI 1.93 to 4.25)). This performance was at a similar level to the worst performing control from both samples ($N = 35$), who was a TD control (who scored 14).

There were additional noteworthy results from the confabulating participants. P2 performed poorly on the Brixton Spatial Anticipation Test (compared with TD controls: two-tailed probability = 0.015, effect size 2.795 (95% CI 1.747 to 3.826)). This performance was not significantly poorer than the ASC controls, but was low within the range of them (at approximately the 11th percentile of the ASC controls). P4 also performed very poorly on the Brixton test, being significantly impaired relative to both the TD controls and the ASC controls (compared with TD controls, two-tailed probability = 0.0001, effect size 5.103 (95% CI 3.336 to 6.857); compared with ASC controls, two-tailed probability = 0.022, effect size 2.606 (95% CI 1.584 to 3.609)). However, there was one outlier in the ASC control group ($N= 17$) who performed at a comparable level (34 errors).

On the Cognitive Estimates test, P2 performed significantly less well than either the TD and ASC control groups in terms of both total score and very extreme estimates (total score: comparison with TD controls, two-tailed probability = 0.021, effect size -2.61 (95% CI -3.59 to -1.62); comparison with ASC controls, two-tailed probability = 0.0041, effect size -3.442 (95% CI -4.71 to -2.16); Very extreme answers, compared with TD controls two-tailed probability = 0.015, effect size 2.77 (95% CI 1.73 to 3.79); compared with ASC controls, two-tailed probability = 0.009, effect size 3.06 (1.90 to 4.21)). None of the ASC or TD groups performed at a level comparable to this (worst performances in the TD group for total score was 10, and within the ASC control group 8, compared with P2's score of 3 (lower scores are worse); for very extreme estimates, poorest TD performance was 5, and for the ASC controls 6, compared with P2's 7 (higher score is worse)). As an aside, it might be worth noting that one of the two ASC controls who produced an unusual number of confabulations to the

Confabulation Elicitation Questionnaire was also the worst performing control participant (of either the TDs or ASC control groups) on the Cognitive Estimates Test.

DISCUSSION

We have described four ASC participants who have shown confabulatory-type phenomena in everyday life. P1 produces florid spontaneous confabulations of the fantastic type. P2 is also a spontaneous confabulator, but the content of them is less fantastic. P3's confabulations are more content specific and fixed, and revolve principally around one topic, and have a paramnesic quality. The relationship between confabulation and paramnesia is a complex one (see Burgess, Baxter, Rose and Alderman, 1996), but the distinction between the two is often made in relation to the breadth and fixedness of the belief held, often with justification for the false beliefs readily proffered. P4 is not a florid confabulator. The content of her confabulations is not grossly fantastic, is not fixed, nor are they content-specific. These four confabulators therefore show different confabulatory phenomena. However, confabulation is not a unitary phenomenon. Rather, it is a term that describes a family of memory errors, and similar forms can be found in the confabulation literature from neurological patients.

Considering the psychometric test findings, three of the four participants have shown impairments on at least some of the tests given to them. The participant whose confabulatory symptoms seem most fixed and circumscribed, almost paramnesic in nature, was also the participant who passed all the experimental tests. The most obvious aspect of these children's presentation however is that the confabulations are largely not occurring in the context of a marked amnesia. P4 was somewhat weak at the word list learning recall and recognition measure relative to the TD controls. But three of the ASC controls actually performed worse (and another performed at the same level), and P4 otherwise performed well at conventional memory tests. The three other confabulators were unimpaired on any of the verbal or visual episodic memory tasks. This presentation is somewhat different from that which typically occurs in neurological patients, where at least some degree of weakness on tests of episodic memory is the most common (although not inevitable) presentation. However, in other ways, the presentations of the confabulators here share features with what is sometimes seen in adult confabulators with acquired disorders. The most florid confabulator (P1) shows weakness on a range of source and context memory tests. P1 also failed one of the executive tasks administered (the Hayling Test), and showed provoked confabulations in response to the questions asked of him in the testing session, which is a type of memory phenomenon often referred to as "momentary confabulation" (Berlyne, 1972). P1 additionally produced two spontaneous episodic confabulations during the testing session. P2 also

showed momentary confabulation, and failed all three of the tests of executive function relative to the TD controls, and two of them relative to the ASC controls. P3 did not fail any of the cognitive tests and did not show momentary or provoked confabulation. P4 however showed a quite different pattern, being weak on the recall and recognition memory tasks, failing aspects of the context memory tests, and showing performance on the Brixton Test that was significantly weaker than either the TD or ASC controls. P4 also showed weak Hayling Test performance, as did P1, P2, and like the ASC control group as a whole.

It is worth noting that as a group, the non-confabulating ASC controls were significantly poorer than the TDs at the recall and recognition episodic memory test. They were not however significantly poorer at the reality monitoring and context memory tests. By contrast, two of the four confabulators failed at least one reality or context memory measure and only one failed the recall and recognition memory test. Of course, the groups in this study are too small to draw firm conclusions from these results. However, in future studies it might be worth considering whether the relationship between amnesia and confabulation may be even less strong in ASC participants than in neurological patients. We can at least say that it appears from these data that the presence of confabulation in ASC participants need not indicate, or result from, a marked amnesia.

It is *prima facie* striking, as mentioned above, that three of the confabulators (P1, P2 and P4) performed poorly on the timed submeasures of the Hayling task (Hayling time B and Hayling time B-A). (One was also poor in terms of Hayling B errors.) However, performance was also poor on these timed measures in the ASC control group compared to the TD control group. This supports previous findings of executive dysfunction in autistic individuals, and specifically, the sensitivity of the Hayling Test in detecting them in this group (White, Burgess & Hill, 2009; Towgood et al, 2009). The Cognitive Estimates task only elicited remarkable results in P2. One process this task specifically examines is monitoring plausibility of responses. Accordingly, poor results might be most expected in those who produce bizarre confabulations. Since none of the subjects had a history of particularly bizarre confabulations, and no confabulations elicited by the confabulation battery were felt to be particularly bizarre, the significance of this pattern on the Cognitive Estimates test is perhaps unclear at this stage.

Indeed, this raises the issue of causality more generally. It is possible, as with many studies that have used similar experimental tasks in neurological patients, that the test results we have observed are unrelated to their confabulation. In other words, that they are epiphenomenal. This seems *prima facie* less plausible where the confabulating participants have performed more poorly than the age- and IQ-matched ASC controls. And in any case, there may be a less direct relation, but one which may be worthy of note nevertheless. For instance, the poor test performances may indicate a predisposing factor, or be an indicator for, or mediating variable for, a causal one. Furthermore, the relation between the diagnosis of autism and the presence of confabulation is no

less certain. More people with autism do not confabulate to an abnormal degree than do. And as far as we are aware, the prevalence of confabulation across many other populations is not well established. So it might be possible that confabulation is no more common in populations of people with a diagnosis of autism than in some others. It is not possible on the present evidence to speak to these complex issues. Overall, the situation of linking a particular symptom to a particular cause is far less easy than for neurological patients with acquired deficits, where they show a behaviour which they did not pre-morbidly. However, the statistical plausibility of such causal matters might perhaps be investigated in much larger ASC (and other developmental conditions) samples, as has been conducted for many other behaviours noticed in people with ASC. However, the variation in presentations found between the confabulating participants argues against a single underlying mechanism of all confabulation in people who show ASC characteristics. In this way, the variety of confabulatory phenomena mirrors the wide variety in patterns of neuropsychological abilities that are found in ASC (Towgood et al, 2009).

There are of course several limitations to this current multiple-case study, some of which suggest ways forward for future research. Although the confabulators and the ASC control group were well-matched by age, ideally they would also have been perfectly matched by sex, IQ (perhaps) and ADOS score. Another limitation is the possible inclusion of undetected confabulators in the comparison groups. Given that every person shows at least some degree of confabulation in everyday life (Burgess & Shallice, 1996b), because our memory systems are highly fallible, it is not clear what exclusion and inclusion criteria should be used. So we made the simplifying assumption that participants were assumed not to show unusual degrees of confabulation unless there were specific reports of this behaviour. However, some formal measure on which selection could be made would have been desirable. It would also have been preferable for the experimenter to have been blind as to group membership, but this was not practically possible for this study. Another possibility for future experimentation would be to present the confabulation elicitation procedure more than once. This might permit examination of the stability of confabulations, and may perhaps elicit differences between neurological and autistic confabulation, since the latter seems not to occur with the same frequency of association with amnesia.

There are some experimental design considerations that are specific to testing an ASC population, or at least that are of more frequent concern than with e.g. neurological patients, and these need to be considered in future experimental designs. For instance, in the reality-monitoring task participants entered responses using computer keys labelled with a pair of eyes for “seen” and a thought-bubble for “imagined”. Testing an ASC population raises the possibility that poor performance might be secondary to difficulty with understanding the representation of thoughts using bubbles, pursuant to a deficit in mentalising (as predicted by Wellman et al, 1996). A counter-argument is that it has been demonstrated that at least some autistic children can understand

thought-bubbles as a way of representing thoughts (Parsons and Mitchell, 1999). In order to try to guard against such a possible artefact, in this experiment participants were allowed to decide for themselves which button represented thought. This ensured that even if a subject did not appreciate the meaning of thought-bubbles, they could respond appropriately by a process of elimination (i.e. since they had two options to choose from). They also practised responding to example questions. This kind of consideration may be one way in which straightforward use of procedures taken from the neurological literature might yield unexpected results unless the special demands of testing this population are considered. Another possible source of error in the test phase was the use of words as cues with no associated picture. Considering the typically delayed language development in ASC, different results may have been elicited if the words were presented verbally, and we will consider this as an addition in future investigations.

A further consideration is whether the executive function deficits here may not be due to executive dysfunction *per se*, but rather due to deficits in social cognition. It has been suggested that individuals with autism may struggle with any task that requires an implicit understanding of the experimenter's expectations for the task, due to their difficulties representing another's mental perspective (White, 2013). For instance, there is evidence of improved performance in people with ASC when they use a computer as opposed to a human examiner on the Wisconsin Card Sorting Test (Ozonoff et al 1995), and tower tasks (Kenworthy, Yerys, Guterman and Wallace, 2008). It would have been preferable to administer the tasks through a computer. For this reason, future experimentation might consider computerised presentation. However, the inclusion of an ASC comparison group for the confabulators goes some way to address this issue. A further consideration is whether the pattern of results would have been different with other tasks of executive function. It has been suggested that people with ASC may perform well on many tests of executive function, apart from the less structured, "open-ended tests" (Kenworthy et al, 2008; White et al, 2009). The Hayling, Brixton, and Cognitive Estimates tests were all chosen for inclusion in this study because all can provide a measure of "bizarre" responding, and there is a clear a priori theoretical relationship with fantastic confabulation. However, if fantastic confabulation is not generally a feature of ASC confabulatory phenomena, then alternative tests of executive function might be logical choices.

A key question for any consideration of confabulation concerns the motivation or cause for the confabulations. Although the confabulatory phenomena, and the patterns of psychometric test performances presented here share features with confabulators shown by people with acquired neurological conditions, it would be premature to draw too strong a conclusion about parallels between the two populations. Neurological patient confabulators are unaware that they are confabulating, and there is no intent to deceive the listener. It is possible that in contrast, the autistic confabulators are fully aware that they are giving invented responses and accounts (of course, in this case we would not formally classify these as confabulations). Perhaps these children do not

appreciate that such behaviour is socially unacceptable, or perhaps they do not even perceive that their audience is not being deceived. Pretence and deception is so prevalent in normal social relations, and in entertainment (films, stories etc), that those who struggle with social convention, and may have limited ability to intuit about what other people are thinking, may understand differently the significance of producing accounts that are inaccurate. We also cannot rule out the possibility that at least some of these confabulations were motivated by gain in some way, emotional or otherwise, knowingly or otherwise. The issue of “motivated confabulation” (e.g. Fotopoulou, Conway and Solms, 2007) within neurological patients is an interesting topic, and one which is the subject of much current debate. It might be useful to consider this debate within the current population also. Certainly, it is clear that being an unreliable witness is not a universal ASC characteristic (McCrory, Henry and Happé, 2007). But perhaps there is a small subsection of the ASC population for whom this is a more prevalent feature. Clearly, more research is needed into the forms, the causes, and the neuroscientific features of this interesting symptom, and this study seems to suggest that applying the methods that have historically been so instructive in the investigation of confabulation might be one way forward.

Acknowledgements

Our thanks are given to Professor David Skuse and his team at the UCL Institute of Child Health / Great Ormond Street Hospital for Children who enabled contact with participants with autism. We would also like to thank Gil Gonen-Yacovi, Roland Benoit, Sam Gilbert, Martha Turner, Rhonda Booth, Wei Chu and Devora Spitzer for their assistance at various stages of the project, and Petra Mäkelä for creating an early version of the confabulation elicitation questionnaire.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions statement

Paul Burgess designed the experiment, contributed to participant interviewing, analysed the data, and co-wrote the paper. Will Mandy organised and recruited participants, interviewed them, supplied clinical data and advised on matters relating to autism, and helped with the write-up. David Spitzer created stimulus materials, tested participants, collected and organised the data, and co-wrote the paper. Sarah White organised and recruited participants, interviewed them, collected and supplied background psychometric data, provided academic advice on matters relating to autism, and helped with the write-up.

References

- Aleman, A., Böcker, K. B. E., Hijman, R., de Haan, E. H. F., & Kahn, R.S. (2003). Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research, 64*, 175-185.
- Asperger, H. (1944). Die „Autistischen Psychopathen“ im Kindesalter. *Archiv f. Psychiatrie* (1944) 117: 76. doi:10.1007/BF01837709.
- Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1994). *Doors and People*. Bury St Edmunds, UK: Thames Valley Test Company.
- Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7*, 212-230.
- Berlyne, N. (1972). Confabulation. *British Journal of Psychiatry, 120*, 31-39.
- Beversdorf, D. Q., Smith, B. W., Crucian, G. P., Anderson, J. M., Keillor, J. M., Barrett, A. M., Hughes, J. D., Felopulos, G. J., Bauman, M. L., Nadeau, S. E., & Heilman, K. M. (2000). Increased discrimination of “false memories” in autistic spectrum disorder. *Proceedings of the National Academy of Sciences, 97*, 8734-9737.
- Bowler, D. M., Gardiner, J. M., Grice, S., & Saavalainen, P. (2000). Memory illusions: False recall and recognition in adults with Asperger's syndrome. *Journal of Abnormal Psychology, 109*, 663-672.
- Bruck, M., London, K., Landa, R., & Goodman, J. (2007). Autobiographical memory and suggestibility in children with autistic spectrum disorder. *Development and Psychopathology, 19*, 73-95.
- Burgess, P.W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*, 547-558.
- Burgess, P. W., Baxter, D., Rose, M., & Alderman, N. (1996) Delusional paramnesic misidentification. In: P. W. Halligan & J. C. Marshall (Eds.), *Method in Madness: Case Studies in Neuropsychiatry* (pp. 51-78). Hove, UK: Psychology Press.
- Burgess, P. W., & McNeil, J. E. (1999). Content-specific confabulation. *Cortex, 35*, 163-182.
- Burgess, P. W., & Shallice, T. (1996a). Bizarre responses, rule detection and frontal lobe lesions. *Cortex, 32*, 241-259.
- Burgess, P. W., & Shallice, T. (1996b). Confabulation and the control of recollection. *Memory, 4*, 359-411.
- Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Bury St Edmunds, UK: Thames Valley Test Company.
- Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839-1849.

- Crawford, J.R., & Garthwaite, P.H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, 40, 1196-1208.
- Crawford, J.R. & Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, 12, 482-486.
- Cunningham, J. M., Pliskin, N. H., Cassisi, J. E., Tsang, B., & Rao, S. M. (1997). Relationship between confabulation and measures of memory and executive function. *Journal of Clinical and Experimental Neuropsychology*, 19, 867-877.
- Dab, S., Claes, T., Morais, J., & Shallice, T. (1999). Confabulation with a selective descriptor process impairment. *Cognitive Neuropsychology*, 16, 215-242.
- Dalla Barba, G. (1993). Confabulation: knowledge and recollective experience. *Cognitive Neuropsychology*, 10, 1-20.
- Dalla Barba, G., Capalletti, J. Y., Signorini, M., & Denes, G. (1997). Confabulation: remembering "another" past, planning "another" future. *Neurocase*, 3, 425-436.
- Dalla Barba, G., Cipolotti, L., & Denes, G. (1990). Autobiographical memory loss and confabulation in Korsakoff's syndrome: a case report. *Cortex*, 26, 525-534.
- Dalla Barba, G., & La Corte, V. (2015). A neurophenomenological model for the role of the hippocampus in temporal consciousness. Evidence from confabulation. *Frontiers in Behavioural Neuroscience*, 9, 218.
- Damasio, A. R., Graff-Radford, N. R., Eslinger, P. J., Damasio, H., & Kassell, N. (1985). Amnesia following basal forebrain lesions. *Archives of Neurology* 42:263-271.
- Fotopoulou, A. Conway, M. A., Solms, M. (2007). Confabulation: Motivated reality monitoring, *Neuropsychologia*, 45(10), 2180-2190.
- Farrant, A., Blades, M., & Boucher, J. (1998). Source monitoring by children with autism. *Journal of Autism and Developmental Disorders*, 28, 43-50.
- Frith, U. (1991) *Autism and Asperger Syndrome*. Cambridge, England: Cambridge University Press. ISBN 0-521-38448-6
- Frith U. (2003). *Autism: Explaining the Enigma*. (2nd ed.). Oxford: Blackwell Publishing.
- Gilbert, S. J., Bird, G., Brindley, R., Frith, C. D., & Burgess, P. W. (2008). Atypical recruitment of medial prefrontal cortex in autistic spectrum disorders: An fMRI study of two executive function tasks. *Neuropsychologia*, 46, 2281-2291.
- Gilboa, A., & Moscovitch, M. (2002). The cognitive neuroscience of confabulation: A review and a model. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *Handbook of Memory Disorders* (pp. 315-342). (2nd ed.). London: Wiley.

- Hala, S., Rasmussen, C., & Henderson, A. M. E. (2005). Three types of source monitoring by children with and without autism: the role of executive function. *Journal of Autism and Developmental Disorders, 35*, 75-89.
- Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: Examining profiles across domains and ages. *Brain and Cognition, 61*, 25-39.
- Hermelin, B., & O'Connor, N. (1967). Remembering of words by psychotic and subnormal children. *British Journal of Psychology, 58*, 213-218.
- Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review, 104*, 499-523.
- Johnson, M. K., O'Connor, M., & Cantor, J. (1997). Confabulation, memory deficits and frontal dysfunction. *Brain Cognition, 34*, 189-206.
- Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences, 2*, 137-145.
- Kenworthy, L., Yerys, B. E., Guterma, A., Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review, 18*, 320-338.
- Kopelman, M. D. (1987). Two types of confabulation. *Journal of Neurology, Neurosurgery and Psychiatry, 50*, 482-487.
- Kopelman, M. D. (1999). Varieties of false memory. *Cognitive Neuropsychology, 16*, 197-214.
- Korsakoff, S. S. (1955). Psychic disorder in conjunction with peripheral neuritis. [Victor, M., Yakovlev, P. I., Trans]. *Neurology, 5*, 394-406. (Original work published 1889).
- 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*, 205-223.
- Luna, B., Minshew, N. J., Garver, K. E., Lazar, N. A., Thulborn, K. R., Eddy, W. F., & Sweeney, J. A. (2002). Neocortical system abnormalities in autism: an fMRI study of spatial working memory. *Neurology, 59*, 834-840.
- McCrory, E., Henry, L. A., & Happe, F. (2007). Eye-witness memory and suggestibility in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry, 48*, 482-489.
- McGlynn, S. M., & Schacter, D. L. (1989). Unawareness of deficits in neuropsychological syndromes. *Journal of Clinical and Experimental Neuropsychology, 12*, 143-205.
- Moulin, C. J. A. (2013). Disordered recognition memory: Recollective confabulation. *Cortex, 49*, 1541-1552.

- Moscovitch, M. (1989). Confabulation and the frontal systems: Strategic versus associated retrieval in neuropsychological theories of memory. In: H. L. Roediger III, & F. M. Craik (Eds.), *Varieties of Memory and Consciousness*. Hillsdale, NJ: Erlbaum.
- Nahum, L., Bouzerda-Wahlens, A., Guggisberga, A., Ptaka, R., & Schnider, A. (2012). Forms of confabulation: Dissociations and associations. *Neuropsychologia*, *50*, 2524–2534.
- Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: a social cultural developmental theory. *Psychological Review*, *111*, 486-511.
- Ozonoff, S. (1995). Reliability and validity of the Wisconsin card sorting tests in studies in autism. *Neuropsychology*, *9*, 491-500.
- Parsons, S., & Mitchell, P. (1999). What children with autism understand about thoughts and thought bubbles. *Autism*, *3*, 17-38.
- Russell, J., & Jarrold, C. (1999). Memory for actions in children with autism: Self versus other. *Cognitive Neuropsychiatry*, *4*, 303-331.
- Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews Neuroscience*, *4*, 662-671.
- Schnider, A., Nahum, L., Pignat, J., Leemann, B., Lövblad, K., Wissmeyer, M., & Ptak, R. (2013). Isolated prospective confabulation in Wernicke–Korsakoff syndrome: a case for reality filtering. *Neurocase*, *19*, 90-104.
- Schnider, A., Gutbrod, K., Hess, C. W., & Schroth, G. (1996a). Memory without context: amnesia with confabulations after infarction of the right capsular genu. *Journal of Neurology, Neurosurgery and Psychiatry*, *61*, 186-193.
- Schnider, A., & Ptak, R. (1999). Spontaneous confabulators fail to suppress currently irrelevant memory traces. *Nature Neuroscience*, *2*, 677-681.
- Schnider, A., Ptak, R., Von Daniken, C., & Remonda, L. (2000). Recovery from spontaneous confabulation parallels recovery of temporal confusion in memory. *Neurology*, *55*, 74-83.
- Schnider, A., von Daniken, C., & Gutbrod, K. (1996b). The mechanisms of spontaneous and provoked confabulations. *Brain*, *119*, 1365-1375.
- Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*, 294-303. *Neurology*. 1981 Sep;31(9):1070-6.
- Shapiro, B.E., Alexander, M.P., Gardner, H., and Mercer B. (1981). Mechanisms of confabulation. *Neurology* *31(9)*, 1070-6.
- Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1990) Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychology*, *28*, 803-813.
- Shingakia, H., Paeksoon, P., Keita, K., Muraic, T., & Tsukiurab, T. (2016). Disturbance of time orientation, attention, and verbal memory in amnesic patients with confabulation. *Journal of Clinical and Experimental Neuropsychology*, *38*, 171-182.

- Skuse, D. Warrington, R., Bishop, D., Chowdhury, U., Lau, J. Mandy, W., Place, M. (2004). The Developmental, Dimensional and Diagnostic Interview (3di): A novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(5), 548 – 558.
- Stolzenberg, S., & Pezdek, K. (2013). Interviewing child witnesses: The effect of forced confabulation on event memory. *Journal of Experimental Child Psychology*, 114(1), 77-88.
DOI: [10.1016/j.jecp.2012.09.006](https://doi.org/10.1016/j.jecp.2012.09.006)
- Stuss, D. T., Alexander, M. P., Lieberman, A., & Levine, H. (1978). An extraordinary form of confabulation. *Neurology*, 28, 1166-1172.
- Towgood, K. J., Meuwese, J. D., Gilbert, S. J., Turner, M. S., & Burgess, P. W. (2009). Advantages of the multiple case series approach to the study of cognitive deficits in autism spectrum disorder. *Neuropsychologia*, 47, 2981–2988.
- Turner, M. S., Burgess, P. W., Cipolotti, L., Frith, C. D., Gilbert, S. J., Shallice, T. , & Simons, J. S. (2008a). Confabulation, reality monitoring and the inferior medial PFC: Converging evidence from lesion and functional imaging studies. *Journal of the International Neuropsychological Society*, 14, 67-68.
- Turner, M. S., Cipolotti, L., Yousry, T. A., & Shallice, T. (2008b). Confabulation: damage to a specific inferior medial prefrontal system. *Cortex*, 44, 637-648.
- Turner, M. S., Simons, J. S., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2008c). Distinct roles for lateral and medial rostral prefrontal cortex in source monitoring of perceived and imagined events. *Neuropsychologia*, 46, 1442-1453.
- Ward, J., & Parkin, A. J. (2000). Pathological false recognition and source memory deficits following frontal lobe damage. *Neurocase*, 6, 333-345.
- Wechsler, D. (1991). The Wechsler intelligence scale for children—third edition. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2004) Wechsler Intelligence Scale for Children - Fourth UK Edition (WISC-IV UK). Oxford, U.K.: Pearson Assessment.
- Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children’s understanding of thought bubbles and of thoughts. *Child Development*, 67, 768–788.
- White, S. J. (2013). The Triple I Hypothesis: Taking another(’s) perspective on executive dysfunction in autism. *Journal of Autism and Developmental Disorders*, 43, 114-121.
- White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on ‘open-ended’ executive functions in autism. *Autism Research*, 2, 138-147.

Table 1. Participant characteristics; means (and standard deviations)

	TD controls	ASC controls	P1	P2	P3	P4
N (M:F)	18 (11:7)	17 (13:4)	M	F	F	F
Age (years)	12.8 (1.8)	13.8 (1.7)	14	11	17	11
VIQ ^a	-	93 (15)	96	-	99	77
PIQ ^a	-	94 (11)	88	73	117	98
3DI ^b - soc	-	14.7 (4.6)	14.9	-	11.4	21.7
- comm	-	14.3 (4.7)	16.2	-	15.2	21.2
- rep beh	-	5.4 (2.9)	10	-	1.3	4.8
ADOS ^c social	-	5.4 (3.2)	7	-	-	6
communication	-	2.7 (1.8)	6	-	-	4
Diagnosis	-	5 autism 8 AS 4 ASC	autism	autism	AS	ASC

^a VIQ = Verbal Intelligence Quotient; PIQ = Performance Intelligence Quotient. IQ data were unavailable for three ASC controls.

^b The Developmental, Dimensional and Diagnostic Interview (3DI; Skuse et al, 2004). 3DI data was unavailable for two ASC controls.

^cThe Autism Diagnostic Observation Schedule (Lord et al, 2000). Data unavailable for two ASC controls

Table 2 Confabulation Elicitation Questionnaire: Example questions

Domain	Example question
Current personal semantic	Do you have any cousins? What are their names?
Remote personal semantic	Who was your first teacher at school?
Current personal episodic	Tell me about your most recent holiday.
Remote personal episodic	Tell me about a trip to the zoo you have been on.
Future personal episodic	Where are you going to eat lunch tomorrow?
Current other semantic	Does your mum have any brothers or sisters?
Remote other semantic	Did your mum go to university?
Current general semantic	Who is the current Prime Minister?
Remote general semantic	Who was Princess Diana?
Don't know semantic	Which country did the longest skateboard journey finish in?
Don't know episodic	What was number one in the charts on your fifth birthday?

Table 3 Results; means (and standard deviations) with statistically significant results in bold.

	TD controls	ASC controls	P1	P2	P3	P4
Confabulation Elicitation Questionnaire (N of confabs)	0.8 (1.3)	1.4 (1.9)	5^{t,&}	7^{t,a}	0	1
Reality monitoring						
- errors ¹	3.1 (4.0)	4.4 (5.7)	18^{t,a}	5	1	7
- I->P errors ²	2.4 (4.1)	3.6 (5.7)	16^t	4	1	7
- I/P->N errors ³	3.2 (3.5)	2.4 (2.4)	0	2	2	3
- N->I/P errors ⁴	0.4 (0.9)	2.1 (5.3)	7^t	2	0	1
Temporal context						
- errors ⁵	2.6 (2.2)	4.2 (4.1)	3	4	5	5
- ?->N errors ⁶	1.3 (1.6)	1.3 (1.5)	3	0	0	0
- N->? Errors ⁷	0.2 (0.6)	1.6 (3.6)	2^t	0	0	2^t
Task context						
- errors ⁸	3.1 (2.2)	4.2 (2.2)	2	-	2	9^{t,a}
- ?->N errors ⁹	1.8 (1.7)	2.4 (3.6)	7^t	-	0	3
- N->? Errors ¹⁰	0.3 (0.8)	1.4 (2.6)	1	-	0	0
Episodic memory						
- recall & recognition ¹¹	29.8 (3.1)	24.7 (6.1)^t	25	29	28	21^t
- new->old ¹²	2.9 (2.7)	3.5 (3.2)	7	4	0	1
- doors & shapes ¹³	48.9 (4.7)	51.6 (3.8)	46	48	50	47
Hayling Test ¹⁴						
- time A	7.5 (2.3)	10.1 (5.0)	11.1	10.8	4.3	8.0
- time B	10.2 (2.8)	23.3 (23.0)^t	25.8^t	24.7^t	8.3	24.9^t
- time B-A	2.8 (3.1)	13.2 (21.0)^t	14.7^t	13.9^t	4.0	16.9^t
- errors B	2.3 (3.4)	3.4 (3.1)	6	13^{t,a}	2	7
Brixton Test ¹⁴	13.1 (3.9)	14.5 (7.1)	7	24^t	16	33^{t,a}
Cognitive estimates ¹⁵						
- total [£]	15.8 (4.9)	17.8 (4.3)	13	3^{t,a}	23	12
- very extreme estimates [§]	2.3 (1.7)	2.1 (1.6)	1	7^{t,a}	2	4

^t Comparison with the TD control group significant at $p < .05$ (or stricter criteria, see text). Comparison between single cases and control groups were conducted using the statistical procedures for single case designs developed by Crawford and colleagues (Crawford, Garthwaite and Porter, 2010; Crawford and Garthwaite, 2002; Crawford and Howell, 1998). Comparisons between ASC and TD control groups were conducted using Mann-Whitney U test.

^a Comparison with the ASC control group significant at $p < .05$ (or stricter criteria, see text).

- Data missing.

[§] Higher scores indicate worse performance.

[£] Lower scores indicate poorer performance.

[&] In addition to the 5 confabulations to the questions listed here, P1 additionally produced two spontaneous confabulations during completion of this task.

¹ Total errors where the participant either mistook a perceived stimulus for an imagined one, or vice-versa.

² Errors where the participant specifically mistook an imagined stimulus as one they had perceived.

- ³ Errors where the participant mistook an earlier item (either imagined or perceived) as a new one.
- ⁴ Errors where the participant identified a new item as one they had imagined or perceived before.
- ⁵ Errors where the participant either thought an item from the first list came from the second or vice-versa.
- ⁶ Where the participant thought that a previously presented item was new.
- ⁷ Where the participant thought that a new item had been presented previously.
- ⁸ Where the participant mistook the question that they had been asked about the image.
- ⁹ Where the participant thought that a previously presented item was new.
- ¹⁰ Where the participant thought that a new item had been presented previously.
- ¹¹ Word list learning – combined recall and recognition.
- ¹² Mistaking one of the (new) foils for an item previously presented.
- ¹³ Visual recognition and recall subtests from the Door and People Test (Baddeley et al, 1994).
- ¹⁴ Burgess and Shallice (1997).
- ¹⁵ Happé et al (2006) version of the Shallice and Evans (1978) test.

Captions to Figures

Figure 1: The reality monitoring paradigm.

