Journal of
Experimental Psychopathology

JEP Volume 4 (2013), Issue 1, 20–37
ISSN 2043-8087 / DOI:10.5127/jep.023711

Forced-Choice Tests as Single-Case Experiments in the Differential Diagnosis of Intentional Symptom Distortion

Thomas Merten[a] and Harald Merckelbach[b]

[a] *Vivantes Netzwerk für Gesundheit, Klinikum im Friedrichshain, Berlin, Germany*

[b] *Forensic Psychology Section (FPN), Maastricht University, The Netherlands*

Abstract

Symptom validity testing has become a prolific research field in neuropsychology. Its original and most powerful version is the forced-choice procedure that continues to play an important role because it allows for an experimental approach to test the hypothesis of intentional symptom distortion. When conducting forced-choice tests, below-chance response patterns are considered to be indicative of this type of distortion. In this article, we discuss the rationale behind the forced-choice technique and its historical development. We also present three case vignettes that illustrate the experimental background of forced-choice testing and how it may help to clarify diagnostic issues. The diagnostic considerations in these cases concerned (1) complete memory loss, (2) Ganser syndrome, and (3) dementia. Employing forced-choice methodology, intentional false symptom production could be demonstrated in all three patients. Thus, the cases demonstrate how, in the context of substantial external benefits, forced-choice testing may help to determine whether healthy individuals try to appear psychologically impaired.

Keywords: Forced-choice testing, Malingering, Symptom Validity Test, Forensic neuropsychology, Amnesia, Ganser syndrome

Correspondence to: Thomas Merten, Vivantes Netzwerk für Gesundheit, Klinikum im Friedrichshain, Department of Neurology, Landsberger Allee 49, D - 10249 Berlin, Germany. Email: thomas.merten@vivantes.de

1. Department of Neurology, Landsberger Allee 49, D - 10249 Berlin, Germany

2. Forensic Psychology Section (FPN), Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands

Received 06-Sep-2011; received in revised form 17-Jan-2012; accepted 19-Jan-2012

Table of Contents

Introduction

In the toolbox of neuropsychology, forced-choice symptom validity testing has a special place. With the forced-choice method, the expert is able to assess the intentionality of false symptom reporting. This type of testing can be conceived as a micro-experiment that the expert conducts with the patient as the only participant. The crux of the experiment is that the patient is exposed to a series of trials on which he or she has to choose between correct target items and incorrect distractor items. The nature of the items depends on the symptoms reported by the patient. For example, items may pertain to autobiographical facts when patients report severe memory loss (i.e., amnesia) or they might be facial expressions when patients report emotional numbing (e.g., Morel & Sheperd, 2008). Target and distractor items may be designed specifically for the case at hand or they may be derived from commercially available standardized tests. With a forced-choice set-up, it is possible to determine whether patients endorse target items with a probability that falls below chance level. If so, this can be taken as a strong indication of the willful choosing of incorrect answers (Frederick & Speed, 2007).

Intentional symptom distortion overlaps with the diagnostic labels of malingering, feigning, and factitious disorder (see Table 1). These labels have serious connotations and the consequences of attributing them to a patient might be far-reaching. Therefore, when experts want to test the hypothesis of intentional symptom distortion in a patient, logical stringency and sound, statistically based decision making are of particular importance. This is all the more true when one realizes that clinical judgements about intentional symptom fabrication often refer to unobservable, purely internal mental states, in particular the patient's awareness of the nature of the symptoms and his/her underlying motivation (Table 1).

Table 1: Symptom distortion and how it relates to DSM-IV-TR terminology (American Psychiatric Association, 2000).

	Intentional symptom distortion	**Unintentional symptom distortion**
External Incentives	*Malingering*	
Internal Incentives	*Factitious disorder*	*Somatoform and dissociative disorders*

In the current article, we first briefly describe the historical background of forced-choice symptom validity testing and then summarize the basic knowledge needed to perform this type of assessment. We next illustrate the technique with three case examples from clinical and forensic referral backgrounds. We conclude with a discussion of some basic issues involved in forced-choice testing.

Historical background of symptom validity testing

Grosz and Zimmerman (1965) published the first case study involving a technique similar to the forced-choice format employed today. The authors investigated a middle-aged man who supposedly suffered from hysterical blindness and whose condition had previously been described by Brady and Lind (1961). Brady and Lind had observed that the patient was sensitive to visual cues. More specifically, they used operant conditioning to show that the patient could deliberately control his behaviour depending upon experimentally manipulated visual input.

During the treatment program, Brady and Lind succeeded to reduce the patient's hysterical symptoms. Eventually, the patient confessed that he could, in fact, see. On readmission because of symptom relapse more than three years later, Grosz and Zimmerman (1965) placed the patient in front of an apparatus that could be operated with switches. Each time when an acoustic signal was presented, the patient had to press one of three switches. If he hit the correct one, the buzzer would stop; with an incorrect response, it would continue sounding for another five seconds. On each trial, the target switch was marked with an upright triangle. The probability of hitting the target switch in the absence of visual cuing was 33.3% (i.e., level of random chance performance), but in a series of consecutive experiments, the patient performed significantly below this level, indicating that he was able to process the visual stimuli and to react to them, although not in the way demanded by the experimental instructions.

In their discussion, Grosz and Zimmerman (1965) raised the question whether below chance-level performance is always a sign of intentional underperformance or whether it might at times reflect hysteria (or modern day somatoform and dissociative disorders). They maintained that experts' opinions on this issue depend on subjective judgments. We will return to this point in the concluding paragraph of this paper.

Grosz and Zimmerman's forced-choice procedure was adopted for other types of clinical cases in which there was the suspicion of intentional symptom distortion. For example, some authors used the technique to evaluate patients who claimed hysterical deafness (e.g., Pankratz, Fausti, & Peed, 1975). These later applications of the forced-choice paradigm mostly employed dichotomous targets and distractors (e.g., high vs. low sound; red vs. green light) so that the probability of chance responding was 0.5. This response format is still the one that is most often used today. Subsequent case descriptions were presented by Miller (1968; patient with unilateral functional blindness), Theodor and Mandelcorn (1973; patient who claimed constricted tubular visual field), Pankratz (1979), Pankratz, Binder, and Wilcox (1987), and Greve, Bianchini, and Ameduri (2003; patients who claimed sensory or tactile loss), Haughton, Lewsley, Wilson, and Williams (1979; patients who claimed hearing loss), and Frederick, Carter, and Powel (1995) and Denney (1996; forensic patients who claimed memory loss).

The forced-choice SVT and its rationale

To develop a tailor made forced-choice Symptom Validity Test (SVT) for a specific case, the clinician or researcher first has to identify a target symptom for which at least two qualitatively distinct states can be defined. These may be two different qualities of light (e.g., red vs. green) in cases of claimed blindness, a high vs. a low pitch tone in cases of claimed auditory loss, two distinct touching points for claimed sensory loss, or two types of memory items (e.g., "I was born in Cologne" vs. "I was born in Munchhausen"; see below) in cases of claimed retrograde amnesia. On the basis of such analysis, the experimenter selects two types of stimuli to which the patient has to respond in a binomial way (e.g., "red" versus "green").

Following a pre-established experimental design (Table 2), the experimenter then presents the two stimuli in a long, pseudo-random order. Often 100 trials are administered. Fewer trial numbers have been utilized, for example, in cases of claimed crime-related amnesia. Denney (2007) recommended a minimum of 25 trials.

Table 2: Example of a pre-established, pseudo-randomized design for 100 consecutive stimulus presentations. A and B refer to target and distractor (e.g., A presentation of red light, B green light). X marks the stimulus condition to be presented at the respective trial. The scheme can also be used as a response protocol for tracking wrong responses.

	A	B		A	B		A	B		A	B		A	B
1	X		21	X		41	X		61		X	81		X
2	X		22		X	42	X		62	X		82	X	
3		X	23		X	43	X		63		X	83	X	
4	X		24		X	44		X	64	X		84		X
5	X		25	X		45	X		65		X	85		X
6		X	26		X	46		X	66	X		86	X	
7	X		27		X	47	X		67	X		87	X	
8		X	28	X		48	X		68	X		88		X
9	X		29		X	49		X	69		X	89	X	
10		X	30	X		50	X		70		X	90	X	
11		X	31		X	51		X	71	X		91		X
12		X	32	X		52		X	72	X		92	X	
13		X	33	X		53	X		73	X		93		X
14		X	34		X	54		X	74		X	94	X	
15	X		35	X		55		X	75	X		95		X
16	X		36	X		56		X	76		X	96	X	
17		X	37		X	57	X		77		X	97		X
18		X	38	X		58		X	78		X	98		X
19		X	39		X	59		X	79	X		99	X	
20	X		40		X	60	X		80	X		100	X	

A forced-choice procedure need not to be time consuming. With cooperative patients, a large number of single responses can usually be obtained within a short time. Patients are instructed to decide on each trial which of the two stimuli was presented (e.g., red *or* green light) or which alternative is the correct one (e.g., my place of birth is Munchhausen *or* Cologne). Furthermore, they are told to guess whenever they do not know the correct answer. This would be the case with patients suffering from genuine blindness, deafness, sensory loss or amnesia. In these patients, the probability of responses follows the binomial distribution (Table 3; see for further technical considerations: Morel & Shepard, 2008; Denney, 1996; Cliffe, 1992).

Table 3: Exact probabilities for forced-choice testing, 100 trials with p = .5.

Number of Correct Answers (*k*)	Exact Probability	Cumulative Probability (One-Tailed)	Cumulative Probability (Two-Tailed)
50	.080	.540	1
49	.078	.460	.920
48	.074	.382	.764
47	.067	.309	.617
46	.058	.242	.484
45	.048	.184	.368
44	.039	.136	.271
43	.030	.097	.193
42	.022	.067	.133
41	.016	.044	.089
40	.011	.028	.057
39	.007	.018	.035
38	.004	.010	.021
37	.003	.006	.012
36	.002	.003	.006
35	.001	.002	.004
34	< .001	.001	.002

Using this distribution, response rates compatible with guessing can be identified, as well as response patterns *below* the threshold for pure guessing. Thus, using statistical decision criteria (e.g., 95 or 99 percent), improbable response behaviour – improbable under the hypothesis that the patient suffers from a genuine impairment and therefore has to guess the correct answer – can be determined. For most researchers in the field of symptom validity assessment, response patterns below chance are the best criterion currently available for identifying *deliberate* attempts to feign neurocognitive disorders (e.g., Slick, Sherman, & Iverson, 1999; Frederick & Speed, 2007). The hidden paradox of the forced-choice SVT is that, in order to score below chance on a forced-choice test the patient must have the cognitive capacity of a person who scores above chance. With this in mind, Iverson (2003; p. 169) proposed a standard formulation for how to report such test performance and its meaning to the referral party (e.g., the court):

“The patient scored below chance on a (...) forced-choice procedure, indicating that she knew the correct answer and deliberately chose the incorrect answer. This performance invalidates the entire set of neuropsychological test results.”

Tables of cut-off scores for different trial numbers and target and distractor probabilities can be found in a number of publications (e.g., Pankratz, 1988), but can also be deduced from the binomial tables in, for example, Siegel and Castellan (1988). The experimenter has to decide beforehand whether one- or two-way significance testing is appropriate. This depends, of course, on the specific hypothesis and the degree of prior information allowing for more specific hypothesis testing (see for a detailed discussion: Frederick & Speed, 2007). For example, when other sources of information (*e.g.*, collateral information) strongly suggest that the patient is feigning his symptoms, one-way hypothesis testing of below-threshold response patterns can be applied. Also, the *a priori* defined significance level may vary depending upon the costs of potential false-positive and false-negative decisions (Schmand & Lindeboom, 2005).

The exact probabilities for every specific number of correct scores as well as cumulative probabilities can be calculated using binomial statistics. That is, the significance of deviations from expected performance (i.e. *pure guessing* by the patient) can be determined for any number of correct responses, for varying total numbers of trials, and varying binomial probabilities. For approximations, z statistics can be calculated, using for cumulative probabilities the following formula: $z = ((k-np) + 0.5)/\sqrt{(npq)}$, with n being the number of trials, k the number of correctly endorsed targets, p the probability of a correct response under conditions of random guessing, and $q = 1-p$.

Through the internet, numerous binomial distribution calculators are available (e.g., stattrek.com/tables/binomial.aspx). Statistical software like SPSS, SAS, Microsoft Excel or Open Office can also be employed. More detailed information about the mathematical assumptions underlying forced-choice testing is provided by Frederick and Speed (2007).

Adaptations for assessment of cognitive impairment

Pankratz (1983) first described the forced-choice technique for the assessment of claimed memory deficits. Unlike forced-choice testing of alleged perceptual deficits, it is in such cases not possible to present a series of homogeneous targets and distractors (e.g., red light vs. green light). Rather individual test items have to be constructed. For example, when a patient claims anterograde amnesia, the experimenter may first present 50 different target words during the learning phase. During the test phase that follows, each trial consists of a target word from the learning list (correct answer) and a distractor word that was not on the learning list (incorrect answer).

This approach has been employed by a number of authors of standardized and commercially available tests, such as the Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Word Memory Test (WMT; Green, 2003). The TOMM and a number of other instruments have all target items presented during one block, while test trials with targets and distractors are presented during a separate block. However, several other tests in the cognitive domain employ the traditional format, with an alternating item-by-item presentation of learning and testing. That is, one target stimulus is presented at a time and forced-choice testing with target and distractor follows immediately (e.g., Computerized Assessment of Response Bias, CARB; Allen, Conder, Green, & Cox, 1997; Bremer Symptomvalidierung, BSV; Heubrock & Petermann, 2000). The apparent disadvantage of this approach is that the patient may easily recognize the test as an SVT, precisely because it is simple and the very low degree of item difficulty is obvious.

In cases of suspicious claims of specific cognitive impairment, such as dyslexia, standardized forced-choice testing is recommended and dedicated SVTs have been developed for this purpose, such as the Word Reading Test (WRT; Osmon, Plambeck, Klein, & Mano, 2006). The Emotional Numbing Test (MENT; Morel, 1998) is another example of a standardized forced-choice SVT. It was developed to assess symptom distortion in patients with alleged posttraumatic stress disorder. More specifically, the MENT taps into the naïve belief that the symptom of emotional numbing might be so extreme that it interferes with the ability of recognizing basic face expressions (see for another example of forced-choice testing in this domain: Rosen & Powel, 2003). Forced-choice testing can also be applied to other cognitive domains that may be crucial in a given case. This may be vocabulary, foreign language knowledge, problem solving ability, mathematical abilities or acalculia (see for a forensic case example, Merten & Puhlmann, 2004).

Sometimes, during clinical assessment, the suspicion of intentional symptom distortion may arise in a completely unexpected fashion. In such cases, tailor made forced-choice testing may yield valuable results.

Forced-choice SVTs in the forensic realm

Brandt, Rubinsky, and Lassen (1985) described the case of 64-year old man charged with the murder of his wife. The man claimed total amnesia for the crime. On a specially designed forced-choice SVT, he scored below chance, indicating that his amnesia claim was false. Frederick and co-workers (1995) extended the application of individualized forced-choice SVTs to forensic assessment. More specifically, the authors demonstrated how presenting patients with a series of target and distractors may enable clinicians to evaluate claims of amnesia. The test items that Frederick and co-workers employed were selected from different content areas: personal information, historical facts, and crime-related details. In their three cases, Frederick and associates found response patterns below chance, suggesting that the amnesia complaints of the patients were feigned so as to obtain insurance money or to be declared incompetent to stand trial in a criminal court.

Denney (1996) administered individualized forced-choice SVTs to specifically test crime-related amnesia claimed by perpetrators. Consequently, his pool of items was limited to details of the crime scene that could only be known to the perpetrator or the forensic examiners. From a forensic point of view, the authenticity of crime-related amnesia claims is an important issue, because (1) a considerable proportion of defendants accused or convicted of violent crimes say that they either partially or completely lost their memories of the offence, (e.g., Pyszora, Barker, & Kopelman, 2003; van Oorsouw & Merckelbach, 2010) and (2) the base rate of genuine memory impairments in such cases is probably dwarfed by malingered amnesia claims. For example, Frederick and Denney (1998) estimated that only about one percent of those who claimed amnesia for their crimes and who were seen by these authors, presented with genuine amnesia, mostly caused by drug intoxication (see also Merckelbach & Jelicic, 2007).

More recent case examples of crime-related amnesia claims evaluated with tailor made forced-choice SVTs are described by Frederick and Denney (1998) and Denney (2009). The referral question in these cases usually pertained to competency to stand trial. For instance, in the case presented by Denney (2009), a 52-year-old man was accused of various criminal offences, but he claimed loss of remote memory. His below-chance performance on a forced-choice SVT strongly suggested that he feigned his memory complaints.

A number of laboratory studies have investigated the potential and the limitations of forced-choice SVTs. Typically, undergraduate participants had to commit a mock crime (e.g., stealing a blue envelope containing 15 euros in a pub) and were then instructed to fake complete memory loss for the theft. Next, they were given a forced-choice SVT that consisted of items such as “the envelope containing the money was (1) blue; (2) green” and “the amount of money stolen was (1) 10 euros; (2) 15 euros.” A recurrent finding in these studies is that a non-trivial proportion (40% to 59%) of participants perform below chance level when they are given an SVT (Merckelbach, Hauer, & Rassin, 2002; Jelicic, Merckelbach, & van Bergen, 2004a,b). This relatively high percentage is remarkable because forced-choice SVTs have been criticized for their transparency, the idea being that healthy people have reasonable statistical intuitions and will see through the rationale behind the SVT and control their responses accordingly (i.e., generate correct responses at or above chance level). The laboratory studies demonstrate that this assumption is not correct, at least not for a considerable proportion of students, even students who are familiar with basic statistical concepts. Forced-choice SVTs can therefore be regarded as useful tools for evaluating the veracity of memory complaints in the forensic context. More precisely, they are best viewed as challenge tests: below-chance performance is a red flag, but random or above random performance has no diagnostic meaning.

Case example: Dissociative amnesia

In a small town in Northern Germany, a 20-year old man attended a party. There was little alcohol and no drug consumption. The conversation was mostly about a sailing event that was to start next morning. At midnight, he went out to the open air and returned after a short while, in a confused state, with a minor skin scrape on his forehead, telling his friends that a group of strangers had been outside the building. Immediately after sharing this information with his friends, he asked them who they were. He claimed not to remember his own name. When his parents were alarmed and appeared a short while after, he claimed not to recognize them either. From there on, he consistently maintained that he had lost all knowledge about himself, about other persons, about everything he had ever learned in school or during his professional training as a carpenter.

Subsequent neurological and neuroradiological examination found no clue whatsoever for a brain abnormality. Thus, it was not long before a psychiatric diagnosis was made: dissociative amnesia. Behavioural therapy was started. The psychotherapist also began to teach him basic preschool and early school knowledge, which he eagerly learned.

During the early phase of symptom presentation, the patient denied any emotion when seeing his family; he explained that he had no feeling of familiarity related to them. However, a couple of nights after the incident, the Holy Spirit itself appeared in his dreams, so he would later report, and revealed the identity of his family members. In his dreams, they were “introduced” to him with words like: This is your mother, accept her in your soul; this is your grandfather, accept him in your heart, and so on. Next morning, he knew who they were and he could emotionally relate to them again.

The patient had subscribed to two very attractive insurance schemes that would enable him to live, for the rest of his life, without financial problems, provided he could prove to be unable to work for health reasons. A pile of clinical reports that he had collected testified to the diagnosis of dissociative amnesia. Not a single medical or psychological report expressed a shadow of doubt as to the veracity of the symptoms reported by the patient.

Two years after the incident, a neuropsychological examination was ordered by one of the insurance companies. At that point in time, the patient explained that he was still trying hard to re-learn every single bit of information he had lost. During the examination, he said that he knew the answer to several questions because he had re-learned the pertinent information after the event. In fact, whenever he knew a piece of semantic or personal information, he was very quick to point out where and when he had re-learned it during the previous two years.

His performance on standard neuropsychological tests was unremarkable, with no signs of cognitive dysfunction, with the exception of poor remote memory recall. Likewise, his performance on standard SVTs (Medical Symptom Validity Test, MSVT: Green, 2004; TOMM: Tombaugh, 1996; Reliable Digit Span, RDS: Greiffenstein, Baker, & Gola, 1994; Rey Complex Figure Test and Recognition Trial, RCFT, memory profile analysis and error analysis: Meyers & Meyers, 1995) was completely inconspicuous. Yet, his performance on the Semantic Memory Test (Schmidtke & Vollmer-Schmolck, 1999) suggested impairment. This test contains, among others, 11 items related to fairy tales of the Brothers Grimm. German children listen to these tales at a very early developmental stage and so this information is usually deeply ingrained in their semantic memory. This patient, however, pretended to be unable to answer any fairy tales items. He explained that there had not been an opportunity to re-learn this category of information after the incident.

Three forced-choice SVTs with tailor made targets and distractor items were designed: one for news events in the patient's home region, constructed on the basis of information in, for example, regional newspapers; a second one for personal information, which was based upon reliable third-party information about the patient's biography; and a third test focusing on knowledge about fairy tales. For this domain, it was possible to formulate a large number of test items. Item examples are given in Table 4.

Table 4: Examples of items of Grimm's Fairy Tale Test: Single-case experimental procedure for testing the hypothesis of intentional fabrication of semantic memory impairment. Answers in italics refer to the correct responses.

Test Item	**Answer 1**	**Answer 2**
What was it Little Red Riding Hood took to her grandmother?	*A bottle of wine*	A bottle of juice
Who helped the prince and Cinderella becoming a couple?	*Pigeons*	Mice
Hänsel and Gretel lost their way in the forest. What was the house built from which they found after a while?	Emeralds and rubies	*Gingerbread*
A guy succeeded in making a very sad princess laugh. He carried…	*A golden goose*	A magic flute
There was a princess who could not sleep. What was the cause of it?	*A pea under the mattress*	A pearl under the mattress
Who helped another princess to get back her golden ball which had fallen into a well?	The Lion King	*The Frog Prince*
Where did Snow White live for a while?	*At the seven dwarfs'*	At the seven ravens'

On the first test (historical and local events), the patient had 38 out of 73 targets correct. This chance performance suggested that he had no knowledge whatsoever about the events presented and therefore merely guessed. On the second test (autobiographical facts), he had only 14 out of 38 target items correct. The exact cumulative probability of attaining this score (or a lower one) by mere guessing is .16. On the fairy tales test, he had only 18 targets correct, out of a total of 50 items. The corresponding probability is .033 (Table 5), which is below chance, indicating that he deliberately avoided the correct answers.

Table 5: Exact cumulative probabilities for a 50-item test used for Grimm's Fairy Tale Test in the first case vignette, with p = .5. For hypothesis testing, one-tailed probabilities were used.

Number of Correct Answers (k)	Cumulative Probability (One-Tailed)	Cumulative Probability (Two-Tailed)
25	.556	1
24	.444	.888
23	.336	.672
22	.240	.480
21	.161	.322
20	.101	.203
19	.059	.120
18	.033	.065
17	.016	.033
15	.003	.006
14	.001	.002

Of course, the story of this patient involved a number of inconsistencies that already from the outset suggested a deliberate act of feigning autobiographical amnesia. There was also a well identified external incentive for the patient. Yet, the results of the third forced-choice test were crucial insofar as the response pattern below chance allowed for a classification of the case as “definite malingering of neurocognitive disorder”, according to the diagnostic criteria proposed by Slick et al. (1999; see below).

The literature is replete with case vignettes of dissociative amnesia and sophisticated discussions about how one can conceptualize the memory impairments in these cases (e.g., whether they are manifestation of implicit or explicit processes; e.g., Allen & Iacono, 2001). We would argue that intentional symptom distortion is a primary issue in such cases and that this issue may be addressed with the type of forced-choice SVTs described above (see also Merckelbach & Rasquin, 2001; Merckelbach, Merten, & Lilienfeld, 2011).

Case example: Ganser syndrome

A key feature of the Ganser syndrome is the tendency to give approximate answers to simple questions. For example, when the patient is asked, to calculate 2 + 2, (s)he will typically come up with a near-miss answer like “5”. Additional symptoms are clouding of consciousness, memory defects, and conversion symptoms. While DSM-IV-TR (American Psychiatric Association, 2000) classifies Ganser syndrome as a dissociative condition, some authors have argued that it is related to brain impairments (Dwyer & Reid, 2004). Remarkably enough, few studies have looked at the performance of Ganser syndrome patients on forced-choice SVTs.

A 50-year-old lawyer crashed with his car into a tree. He was admitted to the emergency ward. MRI scans of his brain were unremarkable. After a few days, he was discharged. The patient said he had tremendous difficulties with recollecting both semantic and biographical facts. He claimed that he had become fully dependent on his wife since the accident. His daily pattern purportedly consisted of staring at the television and going out for small walks with his dog. He complained of headaches, lack of concentration, and feelings of exhaustion. The patient said that he would not be able to return to his law firm. He claimed disability insurance benefits. The insurance company referred him to a psychiatrist for an evaluation. This expert noted that the patient systematically reacted with near-miss answers to simple questions. Other symptoms were a clouded consciousness, memory complaints, and an unstable gait. The psychiatrist concluded that the patient suffered from Ganser syndrome. He considered it to be a dissociative reaction to the accident. As the patient said to be unable to remember all sorts of details from his youth (e.g., his school, his friends), the insurance company referred him to us for a psychological evaluation.

We saw the patient three years after the car crash. He told us that his cognitive functioning had not improved during that period, a conclusion that was confirmed by his wife. During the interview, the patient exhibited a tendency to give approximate answers to simple questions (e.g., he said that he was born in 1953 rather than 1954). We gave him standardized screening instruments for malingering, like, for example, the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005), a self-report instrument that taps into the tendency to over-endorse rare and atypical symptoms. On these tests, the patient clearly over-reported symptoms. We then decided to administer a Famous Faces Test (Merckelbach, Peters, Jelicic, Brands, & Smeets, 2006).

The first phase of the test involved 60 trials that consisted of pictures of Dutch politicians, artists, writers, and sportsmen, who were famous in the 1950's through 1990's. Pictures were presented in a random order, each with a duration of four seconds. The patient was not required to give the names of the famous persons, but just had to classify them into four categories (politicians, sportsmen, writers, artists). Formally, chance level for correct classifications is 25%. However, we also gave this first part of the test to 11 German undergraduates who had little knowledge of Dutch society. Their mean percentage of correct classification was 59% (SD = 6.3). This above chance-level performance probably has to do with the fact that photographs of famous people often contain cues as to the background of their reputation (e.g., a typewriter in the case of a well-known novelist). The Ganser patient attained an overall percentage of correct classifications of 33%. This was somewhat above chance level, but it was far below the level of naïve German students. During the second part of our Famous Faces Test, we showed the patient 12 pictures that had been presented in the first run and 12 famous Dutch people that had not been presented earlier. This time, his task was to say whether the pictures were old or new. Here, his performance fell below chance level (i.e., 25% correct against 50%; $p < .05$). The German students attained a mean percentage of correct responses that was well above chance level on this old-new discrimination task (i.e., 83%; SD = 12.3).

We decided to collect more collateral information in the public domain about the Ganser patient. It turned out that during the period that we had seen him, he had been busy as an enthusiastic tennis player, who was participating – not without success – in two large competitions. He also volunteered as an umpire in field hockey. It is difficult to imagine how a patient who fails to recall the day of the week or his own birthday date, is able to serve as an umpire or play a tennis match. Considering all the evidence, we had little doubt that the patient's Ganser syndrome was fabricated, so as to obtain insurance payments.

The case illustrates four points. First, *ad hoc* SVTs like a Famous Faces Test, a Famous Event Test or a Dead/Alive Test (see for examples, Manning, 2002) might be informative when patients claim amnesia for semantic information. Second, the results of such SVTs become more meaningful when they are also administered to controls like – in our case – naïve German students or individuals instructed to fake a memory disorder. Third, the SVTs may be less transparent and sensitive to strategic responding if they include multiple chance levels. Fourth, tailor made and/or standardized SVTs should be part of the test protocol in patients who present with a rare condition (e.g., massive retrograde amnesia after a stressful period, Ganser syndrome after a mild accident, posttraumatic stress disorder after a mild life event). Failing to include such tests makes it impossible to rule out fabricated psychopathology. Ganser syndrome is a case in point: published case studies of this syndrome abound, but almost none of them included an SVT. Thus, it may well be that the clinical literature about this syndrome is heavily biased by undetected malingerers (Merckelbach et al., 2006).

Case example: Alzheimer's disease

A 55-year old woman was referred for neuropsychological assessment by a court, because she had filed a disability claim that was not accepted by the public pension fund. She was born in a rural region of Kosovo. At the age of 18, she had come to Germany to live with her husband. At the age of 45, her husband separated from her. She continued to live on her own. Each year, she spent the summer holidays with her parents in Kosovo. For a number of years, she had worked as an unskilled worker in a machine factory. At the age of 50, she lost her job and never found a new one. She was obliged to live from welfare. She applied for a disability pension from the public pension fund, but none of her ailments, such as high blood pressure, lower back pain, and mild coxarthrosis qualified for that. In the report of the first independent medical examination performed in 2007, the doctor described her as fully oriented, without clinically obvious impairment of memory or attention.

In her endeavour not to give up the case, she later claimed to suffer not only from depression, but also from significant memory impairment. By the end of 2008, she began to visit a psychiatrist, who conducted a Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) on which she scored only 13 points (full score: 30). The psychiatrist diagnosed Alzheimer's disease and prescribed acetylcholinesterase inhibitors. Her brain scans were unremarkable. Neither a comprehensive neuropsychological assessment, nor cerebrospinal liquor analyses were performed.

On the basis of the reports obtained from her psychiatrist, she claimed disability again and sued the pension fund. In 2009, the judge ordered a neurological evaluation. During the court-ordered neurological examination, the patient displayed a number of inconsistencies in her presentation, so much so that the neurologist doubted a genuine cognitive impairment. For example, she was able to give a detailed account of her medical history, including exact dates, diagnoses, and the names of doctors and other medical staff, but she was not able to do a simple retention task and claimed not to remember a single word or number presented to her. On the basis of these clinical observations, the neurologist proposed a neuropsychological assessment.

The interview started with a remarkable incident. The patient was accompanied by a neighbour who claimed to be a compatriot and a friend of hers. The friend explained that the patient practically spoke no German, so he would translate from Albanian language. During this introduction, the patient herself was sitting on her chair passively, without uttering a word, apparently absent minded. When the neuropsychologist addressed her directly, she would shrug her shoulders and say words like "*nothen understands.*" "You see," the neighbour explained, "you cannot communicate with her at all. So, you need me as a translator. I also accompany her when she sees a doctor." The neuropsychologist explained that he would need a professional translator approved by the court. He would have to contact the judge and discuss the point with him. As a consequence, the patient would be asked to come back at a later stage when everything was settled. At this point, the patient became more alert, moved uneasily in her chair, and finally said, almost inaudibly: "*can speak.*" Upon further inquiry, she explained that she thought she could sufficiently speak German to do the assessment without a translator and, indeed, she did.

The patient gave a detailed account of her history, with all the names of doctors and hospitals that had been involved in her case. She also described her memory problems in depth. At the time of the examination, she continued to live independently, did her shopping, continued handling her financial affairs, and so on. She reported how she had recently visited her ill mother in Kosovo, but was fast to add: "It was only for a couple of days, you understand?"

On standardized cognitive tests, she exhibited major difficulties. For example, this time, she attained only 11 points on the MMSE. On both the Digit Span Forward and Backward (Wechsler, 1987), she performed far below average (2nd percentile each). On a simple motor reaction test (Zimmermann & Fimm, 2002), she obtained extremely low scores (< 1st percentile), as she did on the WAIS-III Block Design (Wechsler, 1997), where she obtained an age scaled score of 2. If these test scores would reflect her true level of cognitive functioning, she would have to be diagnosed with advanced dementia. However, such a serious cognitive impairment would have been incompatible with her living independently, doing housework, using the public transport system on her own, visiting her children who lived in another area of the city, and travelling to Kosovo to visit her parents. Furthermore, it would also be grossly inconsistent with her status of being legally and financially competent, and with the way she managed her pension claims.

Two standardized SVTs were administered, the 1-in-5 Test (Gubbay, n.d.; Tydecks, Merten, & Gubbay, 2006) and the Test of Memory Malingering (TOMM, Tombaugh, 1996). The 1-in-5 Test was adapted from the Digit Memory Test (Hiscock & Hiscock, 1989). In contrast to the typical standardized SVT with 50% chance level, this test was designed such that chance performance yields a score of about 80% correct. It consists of three blocks of 12 trials, with the blocks only differing in the delay following the stimulus card: 5, 10, and 15 seconds, respectively. On each trial, a stimulus card containing four numbers is presented for five seconds, with the instruction to the patient to remember the numbers. Following a delay, a response card is presented that contains all four target numbers and one distractor number. The patient is asked to identify *any one* target on the response card that was also on the stimulus card. The resulting probability of choosing a correct answer by mere chance is 80%. The testing principle is illustrated in Figure 1 (with some modifications, for the purpose of test protection).

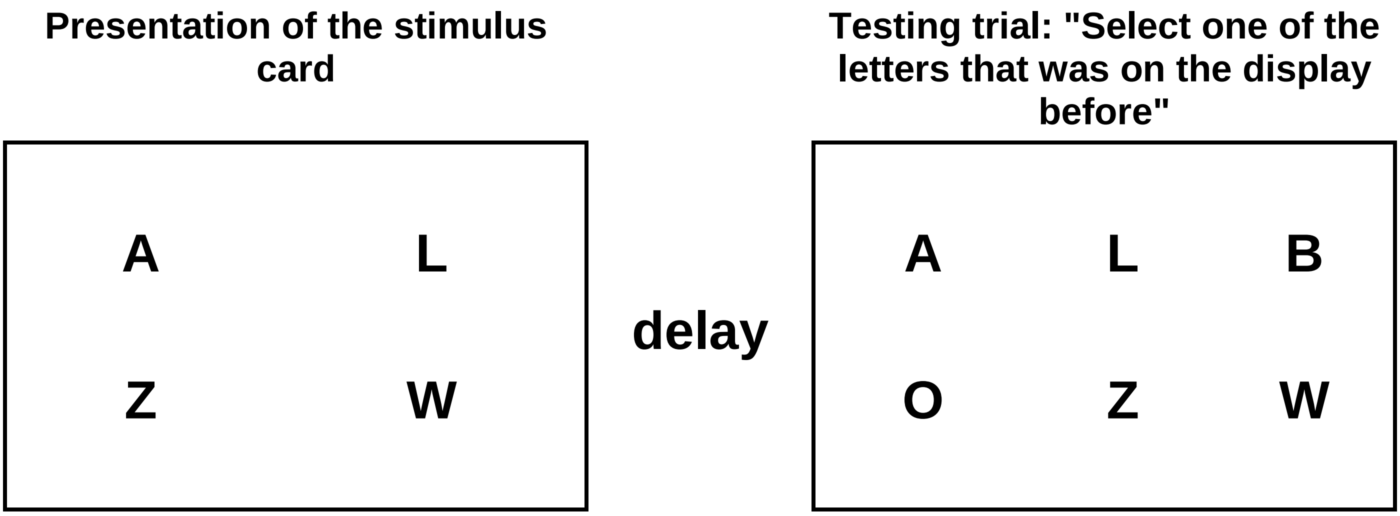

Figure 1: 1-in-5 Test (Gubbay, n.d.): Illustration of the testing principle. For the purpose of this illustration, the test is converted into a 2-in-6 test, using letters instead of numbers. The resulting chance level would be 2/3 in the illustration, while it is 4/5 in the original. The delay between presentation of the stimulus card and the testing is 5 sec for the first set of items, 10 sec for the second, and 15 sec for the third. This variation intends to increase the perceived difficulty.

The TOMM (Tombaugh, 1996) is a standardized SVT which is in wide use in forensic neuropsychology. A total of 50 target stimuli (drawings of common objects) are presented, followed by a forced-choice recognition trial with the target picture and a distractor (chance responding: 50%). Three test blocks are given, each of them containing identical targets, but different distractors. The cognitive load of this test is so low that children in early school age can do the recognition task near to perfection in the second and third test trials (Blaskewitz, Merten, & Kathmann, 2008; Constantinou & McCaffrey, 2003).

The patient scored four, three, and four correct in the three 1-in-5 blocks, respectively. Her total score was therefore 11. With a probability of 80% correct when responding by guessing, this performance reflects an extremely high probability of intentional wrong error production for all three blocks. The post-hoc binomial cumulative probability of scoring 11 or less out of 36 amounts to .000000000193 or 1.93×10^{-10}. In TOMM trial 1, she scored 17 correct (out of 50), which is also below-chance. Therefore, trials 2 and 3 were not given. With reference to the Slick criteria (Slick et al., 1999), the patient was classified as a case of definite malingering, with all the following three criteria being met:

1. Presence of a substantial external incentive;
2. Below-chance performance ($p < .05$) on one or more forced-choice measures of cognitive function;
3. Behaviours meeting criterium 2) are not fully accounted for by psychiatric, neurological, or developmental factors.

This patient is an example of “pseudodementia”. The prefix “pseudo” refers to the fact that the patient presented with a dementia-like condition, but the underlying causes of the claimed cognitive dysfunctions were different from those seen in, for example, Alzheimer's disease. As is true for the Ganser syndrome, “pseudodementia” is probably another condition with a high incidence of malingered symptomatology, although malingering usually is not screened for in such cases. This point has recently been discussed in more detail by Green and Merten (2013).

Conclusions

The diagnostic approach described in this paper is the only one that has gained wide acceptance as a method to uncover intentional response distortion (e.g., Bianchini, Greve, & Glynn, 2005; Slick et al., 1999), so much so that below-chance performance has been dubbed “the smoking gun of intent” (Pankratz & Erickson, 1990). While the methodology of forced-choice SVT offers a unique experimental approach to routine clinical and forensic decision making, there are several problems that remain to be addressed in future research.

One problem is the prior knowledge that patients may have about forced-choice SVTs. Note, for example, that the forced-choice method has found its way into popular TV drama (*Cracker – Mad Woman in the Attic*, directed by Jimmy McGovern, Great Britain 1993). The idiosyncratic response format of alternative choice (i.e., targets versus distractors) can be easily recognized by patients and, consequently, appears to be prone to coaching (e.g., Wetter & Corrigan, 1995; Youngjohn, 1995; Verschuere, Meijer, & Crombez, 2008). Coaching reduces the sensitivity of forced-choice SVTs, and sometimes even reduces it to zero (Verschuere et al., 2008). With the ready availability of information on the internet, we have to expect an ever increasing percentage of patients, claimants, and defendants who are well prepared regarding the methods that will be employed by clinicians to detect potential malingering. They may have been either coached by their attorneys or another third party or found detailed information through effective searching in the internet (Bauer & McCaffrey, 2006; Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002). Different response formats, such as 3-in-5 (Schmand & Lindeboom, 2005) or 1-in-5 (Gubbay, n.d.) may render successful coaching more difficult.

Alternatively, one could add bogus items (i.e., items that have no correct alternative) to the SVT or a “stealth version” (Green, 2004) that is, contrary to the real test, much more difficult. It may well be the case that such strategies interfere with patients’ ability to develop a clever strategy to beat the SVT. Although the case examples above as well as the case studies presented by Denney (1996) show that even with a relatively small number of items a forced-choice SVT might work reasonably well, increasing the number of items will enhance the sensitivity of the test. For example, runs tests conducted on scores of patients who perform in the random range become more powerful when the number of test items increases (see also Cliffe, 1992). Runs tests capitalize on the fact that in general, people find it difficult to generate truly random responses. With run tests, it is possible to determine whether chance level performance is really the result of guessing or whether a certain response strategy (e.g., alternating correct and wrong answers) underlies this type of performance.

Another way of tackling the problem of coaching is the use of built-in or embedded symptom validity indicators (e.g., Suhr & Barrash, 2007) that are more difficult to outsmart, but that usually come along with considerably lower classification accuracy. Such indicators are derived from standard neuropsychological test results and, today, are also conceived as symptom validity tests in the broad sense of the concept (e.g., Greve & Bianchini, 2009). As they follow a different methodological approach, we shall not further discuss them here.

A second major limitation of the forced-choice SVT approach is that below-chance response patterns are expected in a limited number of malingerers because they only occur when response behaviour is grossly distorted (Bickart, Meyer, & Connell, 1991). It is safe to interpret the results of analogue studies with experimental simulators (e.g., Merckelbach et al., 2002) as estimates of the *upper limits of sensitivity*, not the true sensitivity of forced-choice SVTs. After all, in such studies, extremely gross response distortions are induced more often than in a real diagnostic situation where the stakes are usually high. In more sophisticated claimants, this approach is doomed to show lower sensitivity for detecting response bias (see also Greve, Binder, & Bianchini, 2009). Nevertheless, in some populations, the sensitivity of the forced-choice SVT may be low, but still not trivially low. Thus, in a study by Chafetz (2008) about 13% of adults applying for disability status at a Louisiana State agency showed below-chance responding on standardized symptom validity tests. In a sample of claimants with a Turkish migration background tested in the German state of North Rhine-Westphalia, a high percentage (21 out of 44; 48%) of responders suspected of malingering produced below-chance patterns on at least one standardized SVT (Blaskewitz, Merten, & Brockhaus, 2009).

In a recent experimental study with a mock-crime-scenario, Giger, Merten, Merckelbach, and Oswald (2010) compared detection rates of tailor made forced-choice SVTs with standardized symptom validity tests (e.g., the Amsterdam Short Term Memory test; ASTM, Schmand & Lindeboom, 2005). The authors demonstrated that the latter group of methods are superior in detecting false amnesia claims. For example, the sensitivity of the ASTM to detect feigned memory loss was 95%, while that of the individualized SVT was 45%. Thus, whenever possible, commercially available and validated forced-choice SVTs should be used in routine clinical and research testing. This is even more true in the forensic realm. Here, courts often demand that instruments employed by forensic psychologists have been subjected to peer review and are generally accepted by experts. Because of their *ad hoc* character, tailor-made SVTs may have difficulties holding up in court. To a large extent, the legal status of tailor made SVTs resembles that of the so called Guilty Knowledge Test (GKT), which is also based on tailor made items. Some authors have argued that the GKT should be admissible in court because it stands on solid theoretical grounds and enjoys a considerable empirical validity (Ben-Shakhar, Bar-Hillel & Kremnitzer, 2002). We believe that when properly designed and administered, much the same is true for tailor made forced-choice SVTs.

The modest sensitivity of forced-choice SVTs has to do with their transparency, but also with the fact that they tap only into one domain of malingering, usually feigned cognitive abilities. However, malingering is a multidimensional phenomenon and there are malingerers who over-endorse psychiatric symptoms rather than intentionally underperform on cognitive tests. Indeed, correlations between instruments that measure over-endorsement and those that measure underperformance often fall in the 0.10-0.20 range (e.g., Dandachi-FitzGerald, Ponds, Peters, & Merckelbach, 2011; Nelson, Sweet, Berry, Bryant, & Granacher, 2007). Nevertheless, while below-chance performance is just one of a variety of indicators for malingering, it is the red flag that allows for a definite diagnosis of intentional response distortions. Lower degrees of diagnostic certainty, but, at the same time, higher sensitivity for the detection of such distortions can be ascribed to test performance that is in the range of chance responding, but remains below empirically established cutoffs (e.g., for patients after mild traumatic brain injury; see also Frederick & Speed, 2007).

What about the specificity of forced choice SVTs? Is it possible that some patients perform below chance and nevertheless are not making an intentional attempt to bias their responses? Are, in other words, false positives possible with this procedure? Of course, when response patterns are significantly below chance – say at $p = .05$ –, there still is the remote possibility that they reflect pure guessing and, therefore, lack of cognitive abilities. For example, in his sample of individuals with mental retardation, Frederick (2003) found the base rate of below-chance performance on a forced-choice SVT to be on the order of 3%. But apart from such rare fluctuation, it is hard to conceive of circumstances in which patients with genuine cognitive impairments perform below chance. One scenario might be that the test items have not been designed in a proper way. For example, when the incorrect answer options in a memory test have a higher *a priori* plausibility than the correct answer options, it is conceivable that patients who have to guess, but also naïve controls will perform below chance level. This scenario should be ruled out beforehand and it can be done with a form of pilot testing known as the Doob and Kirshenbaum (1973) procedure, which results in a set of unbiased items. As said before, with this procedure, potential target and distractor items of an SVT designed to evaluate a claim of amnesia are pre-tested in healthy, but naïve individuals who are instructed to respond honestly to the test. If they perform below chance, test items should be reformulated.

Tailor made forced-choice SVTs have been applied to patients with genuine neurological conditions. Germane to this are the studies by Weiskrantz (1986) on neurological patients with damage to the projections areas of the occipital lobes. These patients have been found to exhibit blindsight, that is, they perform above chance on visual SVTs, which reflects the operation of implicit perception. Such findings lend further credence to the conclusion that below-chance performance points in the direction of intentional underperformance. It is true that in the past, some studies have found below-chance performance in patients who had been diagnosed with functional blindness or other somatoform disorders (see for example, Bryant & McConkey, 1999). But does that mean that unconscious mechanisms may induce a preference for wrong answer options ultimately leading to unintentional below-chance performance, as some authors have suggested (e.g., Bierley et al., 2001; Drob, Meehan, & Waxman, 2009; for a discussion, see Boone, 2007)? Not at all. With the findings of Weiskrantz (1986) in mind, we would argue that this shows that the diagnostic standards for somatoform disorders in these older studies were low and their samples might have been distorted by malingerers.

Another point to consider here is that people with a genuine somatoform, dissociative, or depressive disorder, may, of course, be uncooperative during a clinical or a forensic examination, and even more so when a substantial secondary gain is at stake. They may invent or exaggerate neurocognitive disorder to make their point clear. This may result in invalid test profiles or even in below-chance performance on forced-choice tests, and yet they do suffer from genuine psychopathology. To be sure, malingering and the presence of genuine psychopathology are not at all mutually exclusive, but the least that one can conclude from below-chance performance is that it invalidates all the other diagnostic information that has been obtained during testing. In any case, rare fluctuations and technical failures aside, below-chance performance on a forced-choice SVT *is* a smoking gun of intentional underperformance.

References

Allen, J. J. B., & Iacono, W. G. (2001). Assessing the validity of amnesia in dissociative identity disorder. *Psychology, Public Policy, and Law*, *7*, 311-344. http://dx.doi.org/10.1037/1076-8971.7.2.311

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB '97 Manual for the Computerized Assessment of Response Bias.* Durham, NC: CogniSyst.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders – DSM-IV-TR* (4th edition, Text Revision). Washington DC: American Psychiatric Association.

Bauer, L., & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the internet: Is test security threatened? *Archives of Clinical Neuropsychology, 21,* 121–126. http://dx.doi.org/10.1016/j.acn.2005.06.010

Ben-Shakhar, G., Bar-Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the Guilty Knowledge Technique in court. *Law and Human Behavior*, *26*, 527-541. http://dx.doi.org/10.1023/A:1020204005730

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal, 5,* 404-417. http://dx.doi.org/10.1016/j.spinee.2004.11.016

Bickart, W. T., Meyer, R. G., & Connell, D. K. (1991). The symptom validity technique as a measure of feigned short-term memory deficit. *American Journal of Forensic Psychology*, *9*, 3-11.

Bierley, R. A., Drake, A. I., Ahmed, S., Date, E. S., Rosner, M., Warden, D., Salazar, A. M., & The Defense and Veterans Head Injury Program (DVHIP) Study Group (2001). Biased responding: A case series demonstrating a relationship between somatic symptoms and impaired recognition memory performance for traumatic brain injured individuals. *Brain Injury, 15,* 697-714. http://dx.doi.org/10.1080/02699050010013671

Blaskewitz, N., Merten, T., & Brockhaus, R. (2009). Detection of suboptimal effort with the Rey Complex Figure Test and Recognition Trial. *Applied Neuropsychology, 16,* 54-61. http://dx.doi.org/10.1080/09084280802644227

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 24,* 379-391. http://dx.doi.org/10.1016/j.acn.2008.01.008

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.). *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29-49). New York: Guilford.

Brady, J. P., & Lind, D. L. (1961). Experimental analysis of hysterical blindness. *Archives of General Psychiatry, 4,* 331-339. http://dx.doi.org/10.1001/archpsyc.1961.01710100011002

Brandt, J., Rubinsky, E., & Lassen, G. (1985). Uncovering malingered amnesia. *Annals of the New York Academy of Sciences, 444,* 502-503. http://dx.doi.org/10.1111/j.1749-6632.1985.tb37625.x

Bryant. R. A., & McConkey, K. M. (1999). Functional blindness: A construction of cognitive and social influences. *Cognitive Neuropsychiatry*, *4*, 227-241. http://dx.doi.org/10.1080/135468099395945

Chafetz, M. D. (2008). Malingering on the Social Security Disability Consultative Exam: Predictors and base rates. *The Clinical Neuropsychologist, 22,* 529-546. http://dx.doi.org/10.1080/13854040701346104

Cliffe, M. J. (1992). Symptom-validity testing of feigned sensory or memory deficits: A further elaboration for subjects who understand the rationale. *British Journal of Clinical Psychology*, *31*, 207-209. http://dx.doi.org/10.1111/j.2044-8260.1992.tb00985.x

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9,* 81-90. http://dx.doi.org/10.1076/chin.9.2.81.14505

Dandachi-FitzGerald, B., Ponds, R. W. H. M., Peters, M. J. V., & Merckelbach, H. (2011). Cognitive underperformance and symptom overreporting in a mixed psychiatric sample. *The Clinical Neuropsychologist, 25,* 812-828. http://dx.doi.org/10.1080/13854046.2011.583280

Denney, R. L. (1996). Symptom validity testing of remote memory in a criminal forensic setting. *Archives of Clinical Neuropsychology, 11,* 589-603.

Denney, R. L. (2007). Assessment of malingering in criminal forensic neuropsychological settings. In K. B. Boone (Ed.). *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 428-452). New York: Guilford.

Denney. R. L. (2009). Evaluating competency to stand trial and sanity in the face of marked amnesia and claimed psychosis. In J. E. Morgan & J. J. Sweet (Eds.). *Neuropsychology of malingering casebook* (pp. 413-427). Hove, UK: Psychology Press.

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups-partial remembering. *Journal of Police Science and Administration*, *1*, 287-293.

Drob, S. L., Meehan, K. B., & Waxman, S. E. (2009). Clinical and conceptual problems in the attribution of malingering in forensic evaluations. *Journal of the American Academy of Psychiatry and Law*, *37*, 98-106.

Dwyer, J., & Reid, S. (2004). Ganser's syndrome. *Lancet*, *364*, 471-473. http://dx.doi.org/10.1016/S0140-6736(04)16772-X

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). " Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research, 12*, 189-198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Frederick, R. I. (2003). *The Validity Indicator Profile* (2nd ed.). Minnetonka, MN: NCS Pearson.

Frederick, R. I., Carter, M., & Powel, J. (1995). Adapting symptom validity testing to evaluate suspicious complaints of amnesia in medicolegal evaluations. *Bulletin of the American Academy of Psychiatry and the Law, 23*, 231-237.

Frederick, R. I., & Denney, R. L. (1998). Mind your "*p*s and *q*s" when using forced-choice recognition tests. *The Clinical Neuropsychologist, 12,* 193-205. http://dx.doi.org/10.1076/clin.12.2.193.2006

Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment, 14,* 3-11. http://dx.doi.org/10.1177/1073191106292009

Giger, P., Merten, T., Merckelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice, 10,* 140-163. http://dx.doi.org/10.1080/15228932.2010.489875

Green, P. (2003). *Green*'s *Word Memory Test. User's Manual.* Edmonton, Canada: Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows. User's Manual.* Edmonton: Green's Publishing.

Green, P., & Merten, T. (2013). Noncredible explanations of noncredible performance on symptom validity tests. In D. A. Carone & S. S. Bush. (Eds.). Mild traumatic brain injury: Symptom validity assessment and malingering (pp. 73-100). New York: Springer.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6,* 218-224. http://dx.doi.org/10.1037/1040-3590.6.3.218

Greve, K. W., & Bianchini, K. J. (2009). Schmerz und Beschwerdenvalidierung [Pain and symptom validity assessment]. In T. Merten & H. Dettenborn (Eds.), *Diagnostik der Beschwerdenvalidität* (pp. 193-229). Berlin: Deutscher Psychologen Verlag.

Greve, K. W., Bianchini, K. J., & Ameduri, C. J. (2003). Use of a forced-choice test of tactile discrimination in the evaluation of functional sensory loss: A report of 3 cases. *Archives of Physical Medicine and Rehabilitation, 84,* 1233-1236. http://dx.doi.org/10.1016/S0003-9993(03)00145-X

Greve, K. M., Binder, L. M., & Bianchini, K .J. (2009). Rates of below chance performance in forced-choice symptom validity tests. *The Clinical Neuropsychologist, 23,* 534-544. http://dx.doi.org/10.1080/13854040802232690

Grosz, H. J., & Zimmerman, J. (1965). Experimental analysis of hysterical blindness. A follow-up report and new experimental data. *Archives of General Psychiatry, 13,* 255-260. http://dx.doi.org/10.1001/archpsyc.1965.01730030061008

Gubbay, J. (n.d.). *Guidelines for 1-in-5 Test.* Unpublished manuscript. Sydney, Australia.

Haughton, P. M., Lewsley, A., Wilson, M., & Williams, R. G. (1979). A forced-choice procedure to detect feigned or exaggerated hearing loss. *British Journal of Audiology, 13,* 135-138. http://dx.doi.org/10.3109/03005367909078888

Heubrock, D., & Petermann, F. (2000). *Testbatterie zur Forensischen Neuropsychologie, TBFN. Testmanual. Neuropsychologische Diagnostik bei Simulationsverdacht* [Forensic Neuropsychological Test Battery. Manual]. Frankfurt, Germany: Swets Test Services.

Hiscock, M., & Hiscock, D. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*, 967-974. http://dx.doi.org/10.1080/01688638908400949

Iverson, G. L. (2003). Detecting malingering in civil forensic evaluations. In A. M. Horton & L. C. Hartlage (Eds.). *Handbook of forensic neuropsychology* (pp. 137-177). New York: Springer.

Jelicic, M., Merckelbach, H., & van Bergen, S. (2004a). Symptom validity testing of feigned amnesia for a mock crime. *Archives of Clinical Neuropsychology, 19,* 525-531. http://dx.doi.org/10.1016/j.acn.2003.07.004

Jelicic, M., Merckelbach, H., & van Bergen, S. (2004b). Symptom validity testing of feigned crime-related amnesia: A simulation study. *Journal of Credibility Assessment and Witness Psychology, 5*, 1-8.

Manning, L. (2002). Focal retrograde amnesia documented with matching anterograde and retrograde procedures. *Neuropsychologia*, *40*, 28-38. http://dx.doi.org/10.1016/S0028-3932(01)00076-8

Merckelbach, H., Hauer, B., & Rassin, E. (2002). Symptom validity testing of feigned dissociative disorder: A simulation study. *Psychology, Crime & Law, 8,* 311-318.

Merckelbach, H., & Jelicic, M. (2007) Amnesia for homicide as a form of malingering. In S. A. Christianson (Ed.). *Offenders' memories of violent crimes* (pp. 165-190). Chichester UK: Wiley.

Merckelbach, H., Merten, T., & Lilienfeld, S. O. (2011). A skeptical look at a remarkable case report of 'overnight' amnesia. Extraordinary symptoms, weak evidence, and a breakdown in peer review. *Skeptical Inquirer, 35* (3), 35-39.

Merckelbach, H., Peters, M., Jelicic, M., Brands, I., & Smeets, T. (2006). Detecting malingering of Ganser-like symptoms with tests: A case study. *Psychiatry and Clinical Neurosciences*, *60*, 636-638. http://dx.doi.org/10.1111/j.1440-1819.2006.01571.x

Merckelbach, H., & Rasquin, S. (2001). Psychobiological research on Dissociative Identity Disorder (DID): A methodological note. *Acta Neuropsychiatrica*, *13*, 68-72.

Merten, T., & Puhlmann, H. U. (2004). Symptomvalidierungstestung (SVT) bei Verdacht auf eine Simulation oder Aggravation neurokognitiver Störungen: ein Fallbericht [Adapting symptom validity testing (SVT) to evaluate suspected malingering of neurocognitive symptoms: A case report]. *Versicherungsmedizin, 56,* 67-71.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial. Professional Manual.* Lutz, FL: Psychological Assessment Resources.

Miller, E. (1968). A note on the visual performance of a subject with unilateral functional blindness. *Behaviour Research and Therapy, 6,* 115-116. http://dx.doi.org/10.1016/0005-7967(68)90051-X

Morel, K. R. (1998). *Manual for the Morel Emotional Numbing Test for posttraumatic stress disorder.* Author.

Morel, K. R., & Sheperd, B. E. (2008). Developing a Symptom Validity Test for posttraumatic stress disorder: Application of the binomial distribution. *Journal of Anxiety Disorders*, *22*, 1297-1302. http://dx.doi.org/10.1016/j.janxdis.2008.01.011

Nelson, N. W., Sweet, J. J., Berry, D. T., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society,* 13, 440-449. http://dx.doi.org/10.1017/S1355617707070373

Osmon, D. C., Plambeck, E., Klein, L., & Mano, Q. (2006). The Word Reading Test of Effort in adult learning disability: a simulation study. *The Clinical Neuropsychologist, 20,* 315-324. http://dx.doi.org/10.1080/13854040590947434

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47*, 409-410. http://dx.doi.org/10.1037/0022-006X.47.2.409

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57,* 367-372. http://dx.doi.org/10.2466/pms.1983.57.2.367

Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 169-192). New York: Guilford.

Pankratz, L., Binder, L. M., & Wilcox, L. M. (1987). Assessment of an exaggerated somatosensory deficit with Symptom Validity Assessment. *Archives of Neurology, 44,* 798. http://dx.doi.org/10.1001/archneur.1987.00520200006007

Pankratz, L., & Erickson, R. D. (1990). Two views of malingering. *The Clinical Neuropsychologist, 4,* 379–389. http://dx.doi.org/10.1080/13854049008401832

Pankratz, L., Fausti, S. A., & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43,* 421-422. http://dx.doi.org/10.1037/h0076722

Pyszora, N. M., Barker, A. F., & Kopelman, M. D. (2003). Amnesia in criminal offences: A study of life sentence prisoners. *Journal of Forensic Psychiatry and Psychology, 14,* 475-490. http://dx.doi.org/10.1080/14789940310001599785

Rosen, G. M. & Powel, J. E. (2003). Use of a Symptom Validity Test in the forensic assessment of posttraumatic stress disorder. *Anxiety Disorders*, *17*, 361-367. http://dx.doi.org/10.1016/S0887-6185(02)00200-1

Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: Misuse of the Internet to assist in avoiding the deception of psychological symptom dissimulation. *Psychological Psychology: Research and Practice, 33,* 294-299. http://dx.doi.org/10.1037/0735-7028.33.3.294

Schmand, B., & Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test –Manual*. Leiden, NL: PITS.

Schmidtke, K., & Vollmer-Schmolck, H. (1999). Autobiographisches Altgedächtnisinterview und semantisches Altgedächtnisinventar [Autobiographical Memory Interview and the Semantic Memory Test]. *Zeitschrift für Neuropsychologie, 10,* 13-23. http://dx.doi.org/10.1024//1016-264X.10.1.13

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Slick, D. J., Sherman E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545-561. http://dx.doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Suhr, J. A., & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In G. J. Larrabee (Ed.). *Assessment of malingered neuropsychological deficits* (pp. 131-170). Oxford: Oxford University Press.

Theodor, L. H., & Mandelcorn, M. S. (1973). Hysterical blindness: A case report and study using a modern psychophysical technique. *Journal of Abnormal Psychology, 82,* 552-553. http://dx.doi.org/10.1037/h0035362

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM).* North Tonawanda, NY: Multi-Health Systems.

Tydecks, S., Merten, T., & Gubbay, J. (2006). The Word Memory Test and the One-in-Five-Test in an analogue study with Russian-speaking participants. *International Journal of Forensic Psychology, 1*, 29-37.

van Oorsouw, K., & Merckelbach, H. (2010). Detecting malingered memory problems in the civil and criminal arena. *Legal and Criminological Psychology*, *15*, 97-114. http://dx.doi.org/10.1348/135532509X451304

Verschuere, B., Meijer, E., & Crombez, G. (2008). Symptom validity testing for the detection of simulated amnesia: Not robust to coaching. *Psychology, Crime, and Law*, *14*, 523-528. http://dx.doi.org/10.1080/10683160801955183

Wechsler, D. (1987). *Wechsler Memory Scale–Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (3rd ed.).* San Antonio, TX: Psychological Corporation.

Weiskrantz, L. (1986). *Blindsight*. New York: Oxford University Press.

Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology: Research and Practice, 26,* 474-477. http://dx.doi.org/10.1037/0735-7028.26.5.474

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology*. Odessa Fl: Psychological Assessment Resources.

Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment, 2,* 279-283. http://dx.doi.org/10.1177/1073191195002003007

Zimmermann, P., & Fimm, B. (2002). *Testbatterie zur Aufmerksamkeitsprüfung (TAP)* [Test for Attentional Performance–TAP]. Version 1.7. Herzogenrath, Germany: Psytest.