SCUOLA DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Psychology

PhD program **Psychology, Linguistics and Cognitive Neuroscience**

Cycle **XXX**

Curriculum in **Social, Cognitive, and Clinical Psychology**

THE BICOCCA SUGGESTIBILITY SCALES: INSTRUMENT DEVELOPMENT AND TESTING INDIVIDUAL DIFFERENCES IN SUGGESTIBILITY IN SCHOOL-AGED CHILDREN.

Surname: **Benedan**

Name: **Laura**

Registration number: **704055**

Tutor: **Dott.ssa Claudia Caprin**

Coordinator: **Prof.ssa Maria Teresa Guasti**

ACADEMIC YEAR 2016/2017

Contents

INTRODUCTION

Over the last decades children have been increasingly involved in forensic context as witnesses. However, research has shown that their testimonies are less accurate and reliable than adults' ones (Ceci & Bruck, 1993). It is widely documented that the accuracy of children's reports hinges largely on the way in which they are interviewed (Gudjonsson, 2003). Children tend to be more accurate when they are elicited with open-ended questions rather than specific questions (Ceci & Bruck, 1993; Chae & Ceci, 2005; Endres, 1997; Geddie, Fradin, & Beer, 2000; Kulkofsky, Wang, & Ceci, 2008), even if their responses often are sketchy and incomplete (Brady, Poole, Warren, & Jones, 1999). Further, the use of specific questions induces children to believe in the existence of a correct answer and generally children avoid answering "I don't know" even when it is an explicit option, preferring to guess (Brady et al., 1999; Roebers & Schneider, 2001a, 2005b). Furthermore, yes-no questions are linked to acquiescence, that is the tendency to respond questions affirmatively regardless of content (Gudjonsson, 1989). Studies have not reported clear response biases in children (Brady et al., 1999; Mehrani, 2010), however, it is likely that children are influenced by the structure of the question and that they tend to reply affirmatively when the question is presented in a positive manner and negatively when a negative question is asked (Mehrani, 2010).

Moreover, several other situational and individual factors may significantly influence the reliability of children's reports. Particularly, in an interrogative situation children might come to accept messages communicated during formal questioning and their subsequent behaviour would be affected. This phenomenon is called "interrogative suggestibility" (Gudjonsson & Clark, 1986) and it deserves particular attention, especially in legal proceedings.

Interrogative suggestibility is a multifaceted concept, which involves social, communication and cognitive skills. Research on child testimony has focused on individual differences underlying children's abilities to provide a complete and accurate

report. Age has been found to be an important predictor of memory performance and vulnerability to suggestibility; however, considerable individual differences also exist between individuals of the same age (Chae & Ceci, 2005). Children's language, their cognitive functioning, and social skills largely impact on their capability to understand questions, to remember events in details and to provide reliable narratives and coherent answers (Powell & Snow, 2007). Indeed, when children are questioned by an adult, cognitive and social factors are intertwined and influence the individual's vulnerability to suggestions (Roebers & Schneider, 2005a). The interplay of cognitive and social factors underlying differences in recall accuracy and suggestibility has been documented in many studies (see Koriat & Goldsmith, 1996; Roebers & Fernandez, 2002; Roebers, Moga, & Schneider, 2001).

Moreover, it has been shown than suggestibility may lead people to falsely believe that they saw details that were only suggested to them, and even cause very rich false memories. Everybody can be tricked by suggestion, especially when memory is uncertain and the technique to induce suggestion is subtle enough (Loftus, 2005). However, individual differences in vulnerability to suggestibility exist. Even though it is still to be proved that suggestibility is a personality trait, this thesis will assume that it is a stable individual tendency that may be measured with a test apt to detect the susceptibility to external misleading influences (Endres, 1997; Gudjonsson, 2003).

The Bicocca Suggestibility Scales (BISS) is a test-battery specifically created for Italian elementary school children to assess children's vulnerability to suggestibility. Children are required to listen to a short story and immediately recall everything they remember about the story. They are subsequently involved in "distracting" tasks for about 20 minutes. Then, they are asked several true and leading questions about the story. Independently of the correctness of their answers, they receive negative feedback and all the questions are proposed again. Lastly, they are required to recall again everything they remember about the story to detect the presence of the post-event misinformation effect.

The first aim of the present thesis is to contribute to the validation of the Bicocca Suggestibility Scales test battery to have a valid and reliable tool to use with Italian school-aged children. The second aim is to assess the relationship between suggestibility and individual differences in cognitive functioning (i.e. intelligence and executive functions), theory of mind, and temperament.

In Chapter 1 a definition of different types of question will be provided, along with their actual use in recommended interviewing protocols. Furthermore, a brief review of the suggestibility research and the relevant theoretical models of suggestibility will be considered. Then, an overview of the existent instruments to assess interrogative suggestibility will be provided.

Chapter 2 outlines in detail the development and composition of the Bicocca Suggestibility Scales (BISS), while Chapter 3 is aimed to prove whether the BISS is a valid and reliable method of assessing individual differences in children's vulnerability to suggestibility. The psychometric properties of the BISS will be examined through four empirical studies. The purpose of Studies 1 and 2 was to assess the test-retest reliability of the parallel forms of the BISS. In particular, Study 1 examined the reliability of the parallel forms of the BISS as well as their internal consistency. In addition to this, Study 2 examined children's memory and suggestibility performance over a two-year delay, while observing the developmental changes in the accuracy of children reports and in their vulnerability to misleading questions and interpersonal pressure. Study 3 examined the concurrent validity of the BISS by comparing it with the Bonn Test of Statement Suggestibility. Study 4 examined the factor structure of the BISS. After having proved the reliability and validity of the instrument, the following studies addressed the role of individual differences in interrogative suggestibility.

In the last section of the thesis four empirical studies will examine whether different individual factors may predict suggestibility using the BISS. For each study, a

brief overview of the research conducted to date on individual differences in children's suggestibility will be offered in relation to the specific topic of interest.

Chapter 4 will examine the relationship between memory and suggestibility with individual differences in fluid intelligence, working memory and executive functions (Study 5).

Chapter 5 will address the influence of theory of mind on interrogative suggestibility. Theory of mind skill will be assessed through two different tasks detecting the ability to infer mental states from auditory and visual cues, respectively. This topic will be investigated first by considering the individual level only. Indeed, in Study 6 situational conditions will be kept constant, and the relative contribution of theory of mind, working memory and inhibitory control will be assessed. Then, Study 7 will evaluate the joint effect of theory of mind and the role of interviewer behaviour.

In Chapter 6 the relationships between memory, interrogative suggestibility, and temperament in school-age children will be evaluated (Study 8). Eight temperamental traits will be considered: social orientation, activity, positive emotionality, negative emotionality, activity, attention, compliance and shyness.

Finally, in Chapter 7 a general discussion will be offered along with the broader implications of this work, its limitations and directions for future research.

Literature review

1.1. Types of questions in interviews with children

The quality of a testimony depends on the dynamic interaction between memory contents, the witnessed event, and the decision-making processes inherent what the witness intends to report (Mazzoni, 2003). Testimony depends largely on the memory of the witness. While a crucial element of the testimony concerns its reliability, a fundamental element of the memory concerns its accuracy (Koriat & Goldsmith, 1996).

The type of questions asked may affect both the quality and the amount of information recalled. Indeed, it is possible to distinguish three general categories of memory recall: *free recall* refers to a memory retrieval process in response to general or open-ended questions; *cued recall* concerns every prompts recollection with a variety of verbal and nonverbal cues (e.g. answers to "Who", "What", "When", "Where", "Why" and "How" questions[1]); *recognition memory* is a form of cued recall in which the respondent picks among various alternatives, including the correct one (Olafson, 2007). It is then crucial to establish when a question may be considered open-ended or closed. To determine which questions may be defined as open-ended or closed questions, they may be classified according to two dimensions: the degree of elaboration and the degree to which they dictate what specific information needs to be recalled. Open-ended questions encourage elaborate, spontaneous descriptions and details (Saywitz & Snyder, 1996; Sternberg, Lamb, Davies, & Westcott, 2001). Open-ended questions are beneficial for response accuracy and minimise individual differences arising from variability in memory, language and social skills (Agnew & Powell, 2004). Open-ended questions

[1] From here onwards these questions will be referred to as "wh-questions".

usually start with the phrase 'Tell me' or the word 'Describe'. This kind of questions should be asked at the beginning of the interview because they encourage witnesses to provide spontaneously an account of the event in their own words. Subsequent open-ended questions should be based on what the witness has just said in free narrative account (e.g. asking for a description of a person the witness has already mentioned 'What did that person look like?') (Bull, 2010).

Specific questions are those questions that ask the child about a particular concept or detail related to the event and generally elicit a one or two word response (Poole & Lamb, 1998). This definition may be applied to cued-recall questions that focus on a given detail that needs to be recalled, as well as specific yes/no and forced-choice questions. In the present dissertation, wh-questions are considered as specific questions. This choice is in line with the most prominent child interview training teams (e.g. Poole & Lamb, 1998; Powell, 2008), even though not all the guidelines for interviewing children define wh-questions the same way either[2].

Yes/no and forced-choice questions provide the interviewee with a limited number of alternatives to choose from. These questions should be the last resource when interviewing children (Bull, 2010).

Another way to classify questions is according to the information they contain. *Leading questions* communicate expectations and premises to an interviewee. They are suggestive since they presume the existence of, or suggest new and misleading information that have not previously been mentioned or disclosed by the child. In other words, a leading question contains highly specific information that has not been mentioned and could subsequently be incorporated into the child's account of the event. This kind of influence may act when the question is in the form of cued-recall (e.g., "What did he do when he was there?"), forced-choice question (e.g., "Was it red or blue?") or yes/no question (e.g. "Did he take your clothes off?"), whenever a new element is

[2] For instance, Wakefield (2006) defined wh-questions as open-ended questions.

introduced as if it was already established as true. A question may be considered as leading even when it implies that a certain answer is accurate or desired (Poole & Lamb, 1998). The leading message may be conveyed by the syntax of the questions or by the intonation of the interviewer's voice. Additionally, forced-choice questions per-se introduce information because they suggest that one of the provided options is the correct answer (see, Brady, Poole, Warren, & Jones, 1999; Poole & Lindsay, 1995).

Leading questions should be avoided because they could undermine the credibility of the child witness. For instance, Waterman and colleagues (Waterman, Blades, & Spencer, 2001, 2004) showed that children may answer in an accurate way when asked answerable questions, and they can recognise that they do not know the answer to unanswerable wh-questions. However, they tend to provide an answer to closed unanswerable questions.

Leading questions are not the only kind of suggestive influence coming from an interview. Other forms of suggestion are questions repetition and imagination requests. Questions may be repeated both across multiple interviews and within an interview. In the first case, questions repetition would increase the likelihood children acquiesce to suggestion and provide additional details (Fivush & Schwarzmueller, 1995); in the latter case, children are more likely to change their response when the question is repeated (Poole & White, 1991). Moreover, Garven, Wood, Malpass, and Shaw (1998) detected several high-impact child interviewing techniques other than suggestive questions by analysing the transcripts of high-profile case that upset the public opinion in the 1990s: social influence, reinforcement, and removal from direct experience (e.g. inviting speculation or the use of puppets to question children) were found to have a great effect on children's answers.

1.2. A theoretical framework for defining suggestibility

1.2.1. The pioneering work of Binet

In 1900, Alfred Binet published the book "La suggestibilité". This can be considered the first attempt to define the suggestive influences in the waking state, contrasting to the suggestibility of the hypnosis. Binet realised several experiments in a primary school in Paris to test the suggestibility of children. He distinguished between a form of autosuggestion – "l'idée directrice" – from the suggestibility induced by external influences, which reflects the mental obedience of the child to the adult (Binet, 1900). The autosuggestion has been studied with an experiment with lines of increasing length. The child's expectation of a progression in the length of the lines led them in error. However, when Binet questioned them, many children were able to detect the errors. In another experiment, the children were required to identify a line having the same length of a model among a series of lines of progressively increasing length. The comparison could be based on the direct perception or based on memory (Binet & Henri, 1894). After, the children had to draw the line as faithfully as possible. Then, the influence of the interviewer's language on children's responses was tested. So, he used the interviewer's feedback to examine the children's vulnerability to interpersonal pressure. After the children had drawn the line, the interviewer said "Are you sure? It is the other line, isn't it?[3]". More than 70% of school-aged children (7-11-year-old) changed their answer, both in the condition based on direct perception, and in the memory condition. Subsequently, he evaluated the influence of the group on children's answers to misleading questions. This experiment brought Binet to the conclusion that social conformity plays an important role in determining a suggested response.

Binet has been a forerunner in the suggestibility research: many conclusions

[3] "En êtes-vous bien sûr? N'est-ce pas la ligne d'à côté?"

drawn from his work have been resumed, tested and replicated in studies conducted several decades after, showing that are still valid. He concluded that children's mistakes in answering suggestive questions were due to gaps in children's memories, filled in while trying to please the adult (Binet, 1900). These errors were then incorporated into children's memory. Indeed, he observed that there are noticeable individual differences in children's suggestibility. Moreover, he noticed that younger children are more vulnerable to suggestion than older ones (Binet, 1900). Finally, his research on children's suggestibility led him to address the issue of the psychology of testimony (Binet, 1904).

1.2.2. Loftus' studies

The studies of Elizabeth Loftus and colleagues are a milestone for memory research. Their studies on the reconstructive memory process started in the 1970s. They showed people a simulated car accident and asked them how fast the cars were going at the moment of impact, by manipulating a simple word in the question - e.g. asking "how fast were the cars going when they *hit* each other?" rather than "how fast were the cars going when they *smashed* each other?". Their results indicated that the witnesses were more likely to say the cars were going faster and that they saw broken glass in the accident scene even if there was no broken glass at all (Loftus & Palmer, 1974). In a subsequent experiment, the participants watched a simulated accident where a car went through an intersection with a stop sign. After receiving the suggestion that there was a yield sign, many witnesses affirmed that they had saw a yield sign at the intersection, not a stop sign (Loftus, Miller, & Burns, 1978). A similar effect was shown even when changing only the phrasing of the question containing the misinformation (e.g. "Did you see *a*…" vs "Did you see *the*…")(Loftus & Zanni, 1975).

The impairment in memory for the past that arises after exposure to misleading information has been defined as *misinformation effect* (Loftus, 2005). The misinformation

effect is usually assessed with a standard three-stage experimental procedure that involve: (1) the presentation of a story, generally in an audio-visual mode; (2) the provision of misleading information that might be presented in the form of a suggestive questioning or erroneous details from another witness; (3) a check on the effect by a recognition task.

Loftus' work proceeded by deepening the understanding of the influence of the misinformation effect. Indeed, after proving that it was possible to add or modify a detail to an otherwise intact memory, Loftus' team's research was orientated and succeeded in proving how to plant an entire memory for an event that never happened. In the 1990s several techniques were used for creating "rich false memories", i.e. to create in people's mind memory for whole events. For instance, participants were induced to believe of having been lost in a shopping mall when they were children (Loftus & Pickrell, 1995), having met Bugs Bunny at a Disney Resort (Braun, Loftus, & Ellis, 2002), and being attacked by a vicious animal (Porter, Yuille, & Lehman, 1999). In a study carried on in Italy, researchers planted the false memory of having witnessed demonic possession (Mazzoni, Loftus, & Kirsch, 2001). Once embraced the suggested event, people can express these false memories with confidence and detail (Loftus, 2005).

Recently, a study demonstrated that even memories for stressful events are highly vulnerable to modification by exposure to misinformation. The participants of that study were members of the U.S. military subjected to a stressful event while they were undergoing a harrowing training exercise simulating the experience of being prisoners of war. As part of this training experience, these soldiers were interrogated in an aggressive and physically abusive way for 30 minutes. After they received suggestive information insinuating that the interviewer was a different person. Then they had to try to identify their interviewer. Approximately 50% of participants misidentified their interrogator (Morgan, Southwick, Steffian, Hazlett, & Loftus, 2013).

These studies suggest that misinformation may distort or change people's memory, and even memory for events that have "just occurred" are vulnerable to misinformation (Morgan et al., 2013).

The importance of research on the misinformation effect is not only contained when people are questioned in a leading way, but it operates even when a witness talk to other witnesses who might provide some erroneous information. Moreover, media coverage may contaminate people's memory (Loftus, 2005).

1.2.3. The Gudjonsson and Clark's model

Interrogative suggestibility may be defined as *"the extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as the result of which their subsequent behavioural response is affected"* (Gudjonsson & Clark, 1986, p.84).

According to this definition, uncertainty, interpersonal trust and expectations are fundamental prerequisites for interrogative suggestibility. Uncertainty acts whenever a person does not know for sure what the correct answer is, because the memory of the event is not strong enough or the question is ambiguous. Interpersonal trust is important because the person must believe that the intentions of the interviewer are genuine and that there is no attempt to deceive. Furthermore, any form of questioning conveys the expectancy that the person to whom the question is addressed should provide an answer (Gudjonsson, 2003).

Several elements of the situation are necessary so that suggestibility occurs: the nature of the social interaction; the interview, which in turn relies on the witness memory; the suggestive stimulus, i.e. leading questions; the acceptance of the suggestion, which may happen only if the message is credible or plausible; the behavioural response.

The structure of the interview can be suggestive in at least two distinct ways: first, the questions may comprise certain expectations and premises, which implicitly suggest to the child what is the desired answer (Endres, 1997). This type of suggestibility is called "Yield to leading questions" (Gudjonsson, 1983, 1984, 1987b). In this view, suggestive questions are those questions that render the interviewer's expectations explicit or that contains premises. Second, the instructions may influence the child's response by providing negative feedback (Gudjonsson, 1984). Therefore, this model puts emphasis on the role of negative feedback and not only on misleading questions. Negative feedback may be any form of overt or implicit criticism of an interviewee answer. It is a kind of interpersonal pressure that could lead the person to modify his or her own previous answers.

The Gudjonsson and Clark's model of suggestibility considers the dynamic relationship between the individual and the environment. It intends to examine the complex process through which people accept premises and incorrect expectations during a police interrogation, coming to provide a false testimony. During this process, both the interviewee's general cognitive set and the interviewer's expectations play a crucial role (Gudjonsson & Clark, 1986). Despite the great theoretical relevance of this model, it should be remembered that it considers only part of a larger process, and many situational determinants could interfere with the individual's reaction.

1.2.4. Ceci and Bruck's definition of suggestibility

A further definition of suggestibility has been proposed by Ceci and Bruck (1995): in their view, suggestibility would refer to *"the degree to which the encoding, storage, retrieval, and reporting of events can be influenced by a range of internal and external factors"* (p. 44).

This broadened definition of suggestibility brings as element of novelty the fact

that it is possible to accept information but be fully conscious of its divergence from the originally perceived event. Therefore, it distinguishes suggestibility from other phenomena such as acquiescence, compliance and lying.

Furthermore, this new concept of suggestibility does not necessarily involve that yielding to suggestive questions implies the alteration of the underlying memory; indeed, a child may still remember what actually occurred even if he or she chooses not to report it. Motivational reasons are included within the range of possible factors underlying a wrong answer. Ceci and Bruck contemplate the possibility that suggestibility may result when misleading information are encountered either before or after an event. Hence, they highlights the role of pre-existent information, such as expectancies (Ceci & Bruck, 1995) or stereotypes (e.g. Leichtman & Ceci, 1995; Lepore & Sesco, 1994). Finally, suggestibility can result from social as well as cognitive factors. This broader framework inspired many studies attempting to ascertain the impact of social and cognitive factors to delineate the underlying causal mechanisms of children's suggestibility (for a detailed review, see Bruck & Melnyk, 2004).

1.3. Suggestibility: State or trait?

Research on suggestibility has been carried on by two different but complementary approaches: the experimental approach (see Loftus' studies) and the correlational approach (see Gudjonsson's studies). While the first kind of studies referred primarily to the conditions under which leading questions are likely to influence the reports of witnesses, the second one was focused on individual differences in responding to leading questions (Schooler & Loftus, 1986). This latter approach considered suggestibility as a trait, i.e. a stable tendency of the individual to respond in a particular way to a given situation (Endres, 1997; Gudjonsson, 2003).

However, the issue whether suggestibility is a state or a trait has been long debated. First, Aveling and Hargreaves' (1921) study defined suggestibility as an individual trait with an approximately normal distribution, being the majority of the subjects moderately suggestible.

Conversely, Baxter (1990) highlighted the importance of situational factors in determining suggestibility, by arguing that even high correlations between different measures of suggestibility would not provide evidence for the trait hypothesis. This because people may accept suggestions in different situations for different reasons.

Endres (1997) criticised Baxter's point of view, by sustaining that high correlations between different measures of suggestibility would provide strong evidence for the existence of a common trait. He reported as example that whenever the same people from a sample scored consistently high on both algebraic and geometric tasks, this would provide support for the existence of a common mathematical ability. This may be considered the factor underlying individual differences for both task types. According to Endres, both situational and individual factors, and possibly their interaction, are necessary to produce suggestibility effects. Situational factors affecting suggestibility are a weak memory representation of the specific event, uncertainty and high trust in the interviewer (Gudjonsson & Clark, 1986), as well as unrealistic expectations of the interviewee's performance. Individual factors include both cognitive and temperamental components.

To date, little empirical research has directly tested the hypotheses underlying trait theories. This could be examined by assessing temporal stability of an individual person's suggestibility scores in a given context, and individual stability of a person relative to his or her peers in one context.

1.4. Instruments to assess interrogative suggestibility in children

Suggestibility is a complex, multifaceted construct that is influenced by both individual and situational factors. Many external determinant could affect the performance of children witnesses, such as characteristics and behaviour of the interviewer, the type of the questions asked and techniques used, the provision of ground rules. Even the nature of the event being recalled, the time delay between the interview and the event may interfere with the accuracy of children's reports. Although the importance of all these situational factors is undeniable, it is necessary to identify the specific characteristics that determine individual differences in children's vulnerability to suggestive influences, especially within the same age groups. In this regard, researchers have developed standardised scales of suggestibility to assess the individual susceptibility to interrogative suggestibility. These scales are suitable to be used with different age-range groups, from pre-schoolers to adults[4].

1.4.1. The Gudjonsson Suggestibility Scales

The Gudjonsson Suggestibility Scales (Gudjonsson, 1984, 1987a, 1997) are standardised tests of memory and suggestibility. Gudjonsson initially developed the GSS1 (Gudjonsson, 1984) as a research and clinical tool designated to provide a measure of an individual adult's level of suggestibility with the aim of assisting in the assessment of pre-trial criminal cases involving retracted confessions (Gudjonsson, 2003). Later, and in order to provide a means for assessing the test-retest reliability of the GSS1, he developed the GSS-2 (Gudjonsson, 1987a), a parallel form of the original scale. Despite having been created and validated with an adult sample, the GSS-2 is the preferred version to use with children.

[4] Even though many paradigms have been used in the literature about children's false memories, the present chapter will consider only the standardised test to assess individual differences in suggestibility.

The GSSs have been extensively used in forensic research involving adults, adolescent and children (Gudjonsson, 2003). The administration of the GSSs is as follows: a short story is read to the child, who is asked to listen carefully and provide a free recall account of the story. Next, the child is asked 20 questions about the story, provided in a standard order. These questions require the participants to state whether specific details had occurred in response to yes/no questions (15 questions) or forced-choice questions (5 questions).

Fifteen of these questions are misleading since they suggest false information, and the correct answer is "no" (e.g. – GSS1 - "Did the woman's glasses break in the struggle?"). The remaining five questions are used to disguise the real purpose of the scale. They are considered "true" questions, because they refer to elements present in the story and their correct answer is "yes" (e.g. – GSS1 – "Did the woman have a husband called Simon?").

After the 20 questions have been answered, the interviewer firmly provides a predetermined negative feedback ("You have made several errors. It is therefore necessary to go through the questions once more, and this time try to be more accurate."), regardless of the participant's actual performance. Then, all the 20 questions are repeated.

A Free Recall index is obtained by considering each salient detail correctly recalled from the story. This score provides an indication of the child's memory ability, as well as whether the child paid enough attention while listening to the story (Gudjonsson, 2003). It is possible that people produce distortion or fabrication of story elements (Clare & Gudjonsson, 1993; Clare, Gudjonsson, Rutter, & Cross, 1994; Gudjonsson & Clare, 1995). Obvious inaccuracies such as direct changes of names of places and characters and of circumstances are counted and constitute the confabulation score, i.e. the amount of story ideas distorted or new ideas introduced (Smith & Gudjonsson, 1986).

The Yield score is derived by calculating the number of times the child capitulates to suggestive questions— i.e., the child provides an affirmative response to the yes/no questions or chooses one option suggested by the false alternative questions.

It is possible to distinguish between Yield1 and Yield2 by considering the number of times the child accepts the suggestion before and after the negative feedback was administered, respectively.

The Shift score refers to the children's tendency to change their responses after receiving negative feedback. One point is given each time the participant modifies one of the previous answers in response to this negative feedback. The Yield and Shift scores are summed to obtain a Total Suggestibility score, which gives an overall measure of the person's vulnerability to interrogative suggestibility. The GSS manual offers the possibility to ask participants to provide a delayed free recall (e.g., after 50 minutes, or one day or one week later).

The GSS has been created for adults but has been used with children too (Danielsdottir, Sigurgeirsdottir, Einarsdottir, & Haraldsson, 1993; Warren, Hulse-Trotter, & Tubbs, 1991). It has also been used with children with intellectual disabilities (Gudjonsson & Henry, 2003; Henry & Gudjonsson, 2003, 2007) and with children who had experienced neglect and were removed from their homes and placed in the care of others. (Benedan, Powell, Zajac, Lum, & Snow, 2018).

The Italian version of the Gudjonsson Suggestibility Scales has recently been published by Curci and Bianco (2014). The reliability and factorial structure have been evaluated by administering both the GSS1 and the GSS2 to a sample of Italian adults (Bianco & Curci, 2015). The authors concluded their paper by recognising that their Italian sample was particularly sensitive to the forensic content of GSS1, while the neutral content of GSS2 was less able to detect individual differences in interrogative

suggestibility. Moreover, the internal consistency of the scales was questionable[5], particularly for the Shift score.

Some data about the use of the GSS2 with Italian children are presented in a recent study (Gudjonsson, Vagni, Maiorano, & Pajardi, 2016). Moreover, the GSS2 has been used with "control" groups of Italian children compared to Italian children being investigated as suspected victims of sexual abuse (Vagni, Maiorano, Pajardi, & Gudjonsson, 2015) or exposed to life adversities (Curci, Bianco, & Gudjonsson, 2016).

1.4.2. The Bonn Test of Statement Suggestibility

The Bonn Test of Statement Suggestibility (BTSS, Endres, 1997) is a standardised instrument used to measure individual differences in interrogatory suggestibility in children aged from 4 to 10 years. This test includes two parallel versions, each containing a brief story, four coloured figures and 31 questions. The test is usually individually administered to the children and the complete procedure takes about 30 minutes: first, the experimenter reads the story, by showing the figures. After, the experimenter takes away the figures and asks the child to provide a free narrative of the story. There is then a pause of about 15 minutes during which the child is involved in tasks unrelated to the story. Hence, the child answers to 31 questions about the story. Specifically, the structured interviewing protocol is composed by 6[6] or 7[7] "distractor" questions, 9 yes/no questions, 8[8] or 7[9] alternative questions and 8 repeated questions. Therefore, three different kinds of misleading suggestive questions are used: yes/no questions, alternative questions, and

[5] The reported Cronbach's alpha were the following:
GSS1 – Yield1: α = .66; Yield2: α = .71; Shift: α = .58; Total suggestibility: α = .68.
GSS2 – Yield1: α = .72; Yield2: α = .77; Shift: α = .53; Total suggestibility: α = .73.

[6] For the version 1 "Toy duck"

[7] For the version 2 "Roller-skating"

[8] For the version 1 "Toy duck"

[9] For the version 2 "Roller-skating"

repeated questions. For each type of suggestive question, separate test scores are obtained, and their sum is computed to give an overall score of suggestibility. Unlike the other instruments, the repetition of questions is more specific, since it regards only a part of the protocol and the question is reproposed immediately after the first answer to the same question.

The BTSS has been developed for German children and its predictive validity of the BTSS (Swedish translation) has been proved in the Finnilä, Mahlberg, Santtila, Sandnabba, and Niemi's (2003) study with children aged 4-5 years and 7-8 years.

The BTSS has been translated in Italian by Rossi, Benatti, Pesce, Oppo, and Avato (2011) and, to date, it has been used in only one study (Volpini, Melis, Petralia, & Rosenberg, 2016). Internal consistency coefficients are between .65 and .85 for the original German version. Regarding the Italian version, the Kuder-Richardson coefficient of the Yield subscale is .78, while the Cronbach's alpha of the Shift subscale is .77 (Rossi et al., 2011).

1.4.3. The Video Suggestibility Scale for Children

The Video Suggestibility Scale for Children (VSSC, Scullin & Ceci, 2001) has been developed to assess the tendency of pre-schoolers to succumb to suggestive questioning.

The administration of the VSSC implies that children view individually or in small groups a 5-minute video clip about a birthday party. The day after viewing the video, children are invited to provide a free narrative of what they remembered about the video ("Do you remember that video about the birthday party? Could you tell me everything you remember about the birthday party video you saw?"). To elicit further details about events in the video, additional open-ended questions are then asked. The video shows a popular mundane event by using elements that are consistent with children's scripts and

expectations (e.g., acts of opening presents, blowing out candles, eating birthday cake), by introducing however some elements that are unexpected (e.g., a fire alarm, clumsy behaviours and broken toys as a present). Matthew Scullin and Stephen Ceci developed this instrument by preferring a video rather than an audiotape of a story because young children experience difficulty encoding information only presented in the oral form (Scullin & Ceci, 2001). Indeed, encoding information presented simultaneously in both auditory and visual form was considered to be similar to the elaboration of an event that children may witness or experience, by ensuring in the meantime the standardisation of administration.

The amount of details correctly supplied by the child is used as an index of memory performance prior to the biasing information. Each of the relevant information correctly remembered is awarded 1 point (on up to 68 points) and contribute to a baseline measure of memory performance: free recall score. The request of providing a free narrative of the video allows the child to rehearse the story and consolidate the event features to warrant that adequate encoding has taken place and that the child is able to reproduce essential contents of the event. Whether the child is not able to accurately recall essential contents of the event, due to lack of attention or insufficient understanding, the scale is considered inapplicable. Next, children are asked 18 specific yes-no questions (e.g. “Was there a girl named Suzie at the party?”), in a standard order. Fourteen of these questions are considered misleading because refer to false information and require a “no” answer (e.g. “Did the two girls arrive at the party in a bright red car?”). Each affirmative answer to a misleading question is given a score of 1 and they are summed to compute the Yield subscale. Therefore, the Yield score corresponds to the total number of questions to which the child incorrectly responds “yes” prior to negative feedback from the interviewer (maximum = 14). After the first 9 questions, children receive mild negative feedback about their responses ("You missed a few of the questions. Let’s go through them again and see if you can do better this time") and then the questions are

repeated. This process is repeated for the second 9 questions. Any change in answer from the previous trial is given a score of one. The number of items where responses are changed after negative feedback is referred to as the Shift subscale. Therefore, the Shift score corresponds to the child's response to negative feedback, and denotes the number of instances in which the child's response changes when the question is repeated. The maximum score is 18 since the four "true" questions are included.

It is possible to compute a composite suggestibility score by summing the child's Yield and Shift scores, thereby permitting a range from 0 to 32.

1.4.4. Why a new instrument for assessing interrogative suggestibility in children?

The Bicocca Suggestibility Scales (BISS) is a standardised test-battery specifically created for Italian elementary school-aged children. Both the Italian version of BTSS and GSS are a translation of tools created in a different context. As noted by Curci and Bianco (2015), cross-cultural variables might be taken into account when considering the Italian version of GSS, because it was administered in a cultural context different from the one in which the instrument was developed.

There are several advantages and disadvantages in choosing to construct a new test rather than to adapt a test from one language and culture to another (Hambleton & Patsula, 1998). Although the validity and usefulness of these tools is unquestionable, the aim underlying the creation of the BISS was to have a flexible tool that may consider that yes–no questions are formulated variously in different languages. For instance, many studies showed that when children answer to yes-no questions, they are likely to display a bias towards saying "yes" (Fritzley & Lee, 2003; Fritzley, Lindsay, & Lee, 2013). It is not clear whether this propensity is a response bias, rather than a tendency to comply with adults' suggestions. This latter hypothesis has been tested by Mehrani and Peterson (2017), who compared responses to positively and negatively framed questions in a

sample of Iranian children and found that children tend to respond to yes–no questions in the direction implied by the question, i.e. "yes" when questions are positively worded and "no" when questions are negatively worded.

Moreover, the BISS has been specifically though for school-aged children by considering that an instrument created for adults may not be suitable for use with children. To render the test more understandable for children, each story has a male and a female child character, to allow the children the identification with one of the protagonists. The stories present a situation in which the characters might be punished; a scenario easily understandable and engaging, giving greater ecological validity to the tool.

The BISS allows the investigation of three different types of suggestive influences: (1) the mere acceptance of leading and misleading information in answering specific questions; (2) the interpersonal pressure underlying the questions repetition within the same interview; (3) the incorporation of suggestions in the subsequent free recall of the event. This latter measure has not been considered by the other suggestibility tests. It has been introduced to consider, within the same test, the distinction between immediate acceptance of misleading information and later retrieval of the misinformation (Schooler & Loftus, 1993).

Moreover, similarly to the GSS procedure, the BISS administration allows to detect the presence of confabulation in both the immediate and the delayed recall of the story. This phenomenon is supposed to occur when people unconsciously fill in gaps in their memory with imaginary experiences (Clare & Gudjonsson, 1993; Smith & Gudjonsson, 1986). This kind of memory error may be a "distortion", when it relates to a change in details of the story (e.g. to attribute an action to the wrong character, to mistake the recipient of an action, to confuse or exaggerate some actions), or a "fabrication", when it refers to the introduction of new elements in the story (Clare et al., 1994; Gudjonsson & Clare, 1995).

However, unlike the GSS, the BISS is administered verbally with the support of pictures, to provide verbal as well visual information, to help the child in maintaining attention towards the task. Indeed, the presence of pictures would help the child in following the story. Furthermore, two parallel versions have been developed to allow the repeated administration to the same child.

Even though the existing tests include some "true" questions (non-leading questions that refer to information presented in the story), these are used with the only aim to disguise the real purpose of the examination. The BISS comprises a consistent number of "true" specific yes/no and forced-choice questions because, in addition to the advantage for face validity, these questions are meant to be used to obtain a standardised measure of recognition memory. Moreover, the long version of the BISS includes the same number of yes/no and forced-choice questions for both the recognition memory and Yield scales; to limit and keep under control the effect of acquiescence, the true questions are balanced so that the correct answer is "yes" for half of them, while it is "no" for the others

Similarly to the GSS, the BISS administration requires that the interviewer repeats all the questions to the child after a mild negative feedback. The answers to all the repeated questions, both "true" and misleading questions, are then considered to form the Shift score. However, the structure of the questions of the BISS makes it possible to distinguish the Shift score into three subtypes, according to the direction of the change: Negative Shift when the child firstly chose the correct answer but changed his or her answer after the interviewer's negative feedback by selecting one wrong option; Positive Shift: when the child provided a wrong answer during the first administration of the questioning protocol and corrected himself or herself after the interviewer's negative feedback; Neutral Shift: when the child provided a wrong answer during the first administration of the questioning protocol and changed his or her answer after the

interviewer's negative feedback, by selecting the other wrong alternative suggested by the interviewer.

This conceptualization of the Shift score has been made considering the psychometric and conceptual problems that have been associated with the Shift score of the GSS (see Gignac & Powell, 2009; Bianco & Curci, 2015) and the VSSC (see McFarlane & Powell, 2002).

The Bicocca Suggestibility Scales

The Bicocca Suggestibility Scales (BISS) is a test battery comprehensive of two parallel forms, A and B. The scale's format, administration procedure and scoring criteria are identical for the two forms of the test. The only difference between the BISS-A and BISS-B relates to the content of the narrative paragraphs and, consequently, of the questions. Because the test relies on a memory paradigm, two parallel forms have been developed to allow the assessment of the test-retest reliability of each scale. Both the BISS-A and the BISS-B contain a short story equipped with 6 pictures and a questioning protocol that may be presented in a short version (22 items) or in a long version (32 items). Ideally, the short version is used as a screening phase to identify children with an outstanding predisposition to succumb to interrogative suggestibility. The long version should be used as in-depth analysis phase to ascertain the extent of the individual vulnerability to suggestion.

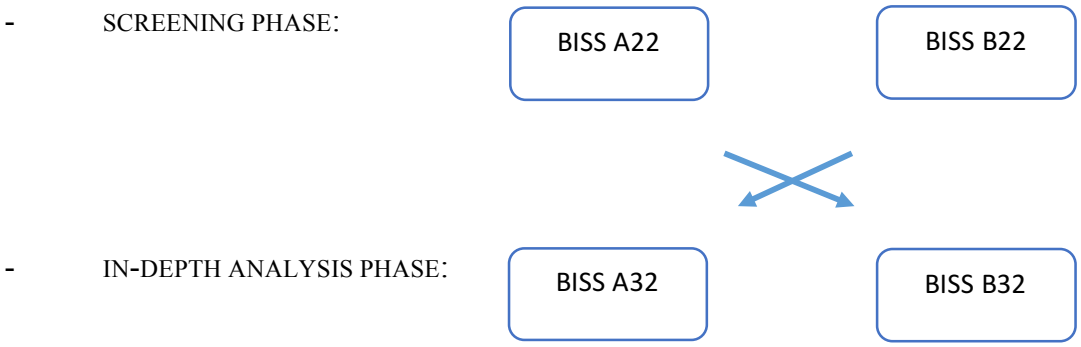
The short versions contain 8 unbiased specific questions (3 yes/no and 5 forced-choice questions) and 14 suggestive questions (6 yes/no and 8 forced-choice questions).

The long versions contain 16 unbiased specific questions (8 yes/no and 8 forced-choice questions) and 16 suggestive questions (8 yes/no and 8 forced-choice questions).

In line with Gudjonsson (1984, 1987a), two different measures of interrogative suggestibility may be detected: *Yield* to misleading questions and *Shift* after negative feedback.

2.1. The assessment

Figure 2.1. The assessment procedure of the Bicocca Suggestibility Scales Battery



2.2. Procedure[10]

The test should be administered individually in a quiet room. Administration takes about 30 minutes and has the following phases:

1. The story is presented by the interviewer (*"I'm going to reading you a story and you have to be very careful to what I'm going to tell you because then I'm going to ask you to repeat it, just as I've read it").* Then, he or she reads aloud the story to the child, while simultaneously showing the pictures. Each story has a male and a female character, to allow the children the identification with one of the protagonists.
2. The interviewer asks the child an open-ended question (*"Tell me the story as I've read it").* This step has two main goals: first, the child can rehearse the content of the story to consolidate the memory trace; second, the free recall of the story is audio recorded and then transcribed to obtain a measure of the memory performance before the provision of suggestive influences. Moreover, it allows to determine if the child attention during the listening of the story and his or her understanding of the narrative was sufficient to

[10] The procedure has been translated in English for the present thesis for a descriptive purpose, while the story, the questions and the instructions have been used in Italian only. The original version of the instrument is shown in Appendix A.

report the essential information about the story. In line with the GSS, the analysis of the transcripts may reveal the presence of confabulation, i.e. spontaneous memory errors committed by the child during the immediate recall of the story. Confabulation may relate to errors about details of the story (e.g. to attribute an action to the wrong character, to mistake the recipient of an action, to confuse or exaggerate some actions), or may refer to completely fictional elements not present in the story and not consistent with the script or additional details that change the course of story (e.g. to attribute emotions to the characters when not mentioned in the story, insert in the narrative actions not originally described or things unsaid). The first type of error is considered a *Distortion*, while the second is considered a *Fabrication*.

3. At this point, there is a pause of about 25 minutes. This interval has the purpose of distracting the child from the content of the story and to weaken the memory trace. The Raven's Coloured Progressive Matrices (Raven, 1998) is the recommended test, since it provides an assessment of the intelligence of the child, without the danger of contaminating the child's memory with new verbal stimuli.

4. After the pause, the questioning phase starts. The interviewer says: *"I'm going to ask you a series of questions about the story that I read before, to understand exactly what you remember. If there are any questions about something you don't remember, don't try to guess, but tell me "I don't know""*. Thus, the interviewer asks the first set of questions.

5. After that, regardless of the actual performance of the child, the interviewer gives a mild negative feedback in a firm tone (*"You've made some mistakes, do you know? Now, I'm going to ask you the same questions I've already ask you. If there are any questions about something you don't remember not to try to guess, but tell me "I don't know").* Then, the interviewer repeats all the questions and after he or she provides an indiscriminate positive feedback to reduce the emotional discomfort of the child.

6. Finally, the interviewer requires a second free narrative of the story, which is audio recorded and subsequently transcribed to obtain a measure of the memory performance after the provision of suggestive influences. The main goal of this delayed free recall is to reveal inconsistencies between the first and the second recall of the story and to detect the presence of the *post-event misinformation effect*. This index refers to the possible presence of elements induced by suggestive questions and then introduced in the delayed narrative of the story. For a detailed analysis, it is possible to identify the specific question which generated the error and to differentiate between the different categories of questions (control Yes/no question; control forced-choice question; suggestive Yes/no question; suggestive forced-choice question).

2.3. Bicocca Suggestibility Scales (BISS) - Form A

This is Anna. / This is Mark. /

They are brother and sister. / Mark / is older than Anna, / but Anna / is taller than Mark. / After / having done their homework, / they are watching the cartoons together. / When the cartoons end / a crime movie / starts. / Anna wants to see the movie / but Mark does not like it. / So, they start fighting / and the remote control falls to the ground / and breaks. / Their mother / hears them arguing / and enters the room. / Mark / picks the remote control up / and hides it / behind his back. / The mother asks: "Why are you quarrelling?" / Anna replies: / "We are not fighting! / We are playing". / Then the mother turns to Mark and asks: / "What are you hiding? / And he shows her the broken remote control. / The mother gets angry / and wants to know who broke it. / Anna says: / Mark threw it / against the wall. / Mark says “That’s not true! / Anna threw it / against the door. / The mother says that, as punishment / Mark / will not be allowed to play / with the play-station / for one week. / Then the mother says: / “The remote control can’t be fixed, / Dad / will have to buy a new one. /

2.3.1. The BISS-A32

First series of questions

1. Did mum get very angry when she saw that the remote control was broken?
2. Did Mark throw the remote control against the wall?
3. Did Anna hide the remote control under the pillow when mum came in?
4. Was the remote control bought by dad?
5. Did the children eat a sandwich with jam or with Nutella?
6. Did Mark say that Anna threw the remote control against the door or against the wall?
7. Did Mark and Anna have their snack?
8. Did mum punish Mark or Anna?
9. Did mum slap Anna or Mark?
10. Did dad get very angry when he knew that the remote control was broken?
11. Was the punishment given by mum not to watch TV or not to play with the Play Station?
12. Did the remote control fall to the ground during the fight?
13. Did mum say that the remote control was broken or that it could be fixed?
14. Who said to mum that they were not fighting but they were playing? Mark or Anna?
15. Did Mark want to see a science fiction movie or a music show?
16. Who will buy the new remote control, mum or dad?
17. Did Mark switch off the TV because he wanted play with the Play Station?
18. Who broke the remote control last week, Mark or Anna?
19. Did the remote control cost a lot of money?
20. Is it true that Mark didn't like the crime movie?
21. Was mum cooking or ironing?

22. Was dad reading in the kitchen?
23. Were the children watching an adventure show or the cartoons?
24. Did Anna and Mark watch the telly after having done their homework?
25. Do Mark and Anna usually do their homework in the kitchen?
26. Does Mark want to be a musician or a football player when he grows up?
27. Is Anna better than Mark at school?
28. Is the television near the door or the window?
29. Does Anna want to be a policewoman when she grows up?
30. Do the children watch TV every afternoon or just sometimes?
31. Was mum doing shopping when the children started fighting for the remote control?
32. Did Anna want to see a crime movie or a documentary about animals?

Repeated questions

1. Mum didn't get very angry when she saw that the remote control was broken, did she?
2. Mark didn't throw the remote control against the wall, did he?
3. Anna didn't hide the remote control under the pillow when mum came in, did she?
4. The remote control wasn't bought by dad, was it?
5. Did the children eat a sandwich with Nutella or with jam?
6. Did Mark say that Anna threw the remote control against the wall or against the door?
7. Mark and Anna didn't have their snack, did they?
8. Did mum punish Anna or Mark?
9. Did mum slap Mark or Anna?
10. Dad didn't get very angry when he knew that the remote control was broken, did he?

11. Was the punishment given by mum not to play with the Play Station or not to watch TV?
12. The remote control didn’t fall to the ground during the fight, did it?
13. Did mum say that the remote control could be fixed or that it was broken?
14. Who said to mum that they were not fighting but they were playing? Anna or Mark?
15. Did Mark want to see a music show or a science fiction movie?
16. Who will buy the new remote control, dad or mum?
17. Mark didn't switch off the TV because he wanted play with the Play Station, did he?
18. Who broke the remote control last week, Anna or Mark?
19. The remote control didn’t cost a lot of money, did it?
20. Is it true that Mark didn't like the crime movie, isn't it?
21. Was mum ironing or cooking?
22. Dad wasn’t reading in the kitchen, was he?
23. Were the children watching the cartoons or an adventure show?
24. Anna and Mark didn’t watch the telly after having done their homework, did they?
25. Mark and Anna don't do their homework in the kitchen, do they?
26. Does Mark want to be a football player or a musician when he grows up?
27. Is Mark better than Anna at school?
28. Is the television near the window or the door?
29. Anna doesn’t want to be a policewoman when she grows up, does she?
30. Do the children rarely watch TV or every afternoon?
31. Mum wasn't doing shopping when the children started fighting for the remote control, was she?
32. Did Anna want to see a documentary about animals or a crime movie?

2.3.2. The BISS-A22

First series of questions

1. Did Anna want to see a crime movie or a documentary about animals?
2. Did the children eat a sandwich with jam or with Nutella?
3. Did mum slap Anna or Mark?
4. Did dad get very angry when he knew that the remote control was broken?
5. Was the punishment given by mum not to watch TV or not to play with the Play Station?
6. Did the remote control fall to the ground during the fight?
7. Did mum say that the remote control was broken or that it could be fixed?
8. Who said to mum that they were not fighting but they were playing? Mark or Anna?
9. Did Mark want to see a science fiction movie or a music show?
10. Who broke the remote control last week, Mark or Anna?
11. Did the remote control cost a lot of money?
12. Is it true that Mark didn't like the crime movie?
13. Was mum cooking or ironing?
14. Was dad reading in the kitchen?
15. Were the children watching an adventure show or the cartoons?
16. Did Anna and Mark watch the telly after having done their homework?
17. Do Mark and Anna usually do their homework in the kitchen?
18. Does Mark want to be a musician or a football player when he grows up?
19. Is Anna better than Mark at school?
20. Is the television near the door or the window?
21. Does Anna want to be a policewoman when she grows up?
22. Do the children watch TV every afternoon or just sometimes?

Repeated questions

1. Did Anna want to see a documentary about animals or a crime movie?
2. Did the children eat a sandwich with Nutella or with jam?
3. Did mum slap Mark or Anna?
4. Dad didn't get very angry when he knew that the remote control was broken, did he?
5. Was the punishment given by mum not to play with the Play Station or not to watch TV?
6. The remote control didn't fall to the ground during the fight, did it?
7. Did mum say that the remote control could be fixed or that it was broken?
8. Who said to mum that they were not fighting but they were playing? Anna or Mark?
9. Did Mark want to see a music show or a science fiction movie?
10. Who broke the remote control last week, Anna or Mark?
11. The remote control didn't cost a lot of money, did it?
12. Is it true that Mark didn't like the crime movie, isn't it?
13. Was mum ironing or cooking?
14. Dad wasn't reading in the kitchen, was he?
15. Were the children watching the cartoons or an adventure show?
16. Anna and Mark didn't watch the telly after having done their homework, did they?
17. Mark and Anna don't do their homework in the kitchen, do they?
18. Does Mark want to be a football player or a musician when he grows up?
19. Is Mark better than Anna at school?
20. Is the television near the window or the door?
21. Anna doesn't want to be a policewoman when she grows up, does she?
22. Do the children rarely watch TV or every afternoon?

2.4. Bicocca Suggestibility Scales (BISS) - Form B

This is Alice. / This is Matthew.

They are cousins. / Alice / has blonde hair / while Matthew / has blue eyes. / They are on the mountain / at grandmother's house. / Grandma / wants to make a cake / but she has run out of eggs. / The grandmother tells the children: / "Go / to the town / to buy some eggs". / The grandmother recommends them / not to take the bike, / because the eggs may break. / The grandmother gives them money. / Children disobey their grandmother / and ride the bike. / They see an ice cream parlour / at the town / and they buy an ice cream / even though they know / that they are not allowed to eat ice cream / before lunch. / Once in the shop / they no longer have enough money / and buy just one egg. / Back home / they give the egg to their grandmother. / Grandma asks them: / "Why did you buy only one egg?" / She adds sadly: / "I can't make you the cake!"/ Alice says: / "Matthew wanted to ride the bike, / he fell / and the eggs broke!" / Matthew gets angry / and says to the grandmother: "That's not true!" / Alice lost the money!"/ The grandmother replies gently: / "I don't believe you. / You are hiding something / and I'll find out what you did!"

2.4.1. The BISS-B32

First series of questions

1. Are Alice and Matthew at grandma's house?
2. Did grandma tell Matthew that she needed 10 eggs?
3. Did grandma tell the children to buy an ice cream?
4. Do the children have to ride along a rise before arriving in town?
5. Is it summer or spring?
6. Did they buy an egg or did the shopkeeper donate it to the children?
7. Did the children stop playing in the woods before going back home?

8. Did the children eat the ice cream before or after lunch?
9. Did grandma wash the clothes or make the beds when the children were in town?
10. Did Alice eat a strawberry ice cream?
11. Did only Alice eat the ice cream or both the children?
12. Are Matthew and Alice cousins?
13. Is grandma's house in the mountain or in the countryside?
14. Did Matthew say that he lost the money or that Alice did?
15. Is the grandma's dog name Toby or Bobby?
16. Is the ice cream parlour near grandma's house or in town?
17. Did grandma tell the children to ride the bike to hurry up?
18. Did grandma want to make a chocolate cake or a cake with jam?
19. Did Matthew and Alice meet their friends when they arrived in town?
20. Did Alice tell grandma that Matthew broke the eggs falling off the bike?
21. Does grandma have a cat or a parrot at home?
22. Are Matthew's and Alice's mother sisters?
23. Did grandma punish Matthew or didn't she believe either of them?
24. Did Matthew and Alice eat the ice cream when they arrived in town?
25. Does Matthew have a younger brother?
26. Is it sunny or is it going to rain?
27. Does the ice cream give Matthew stomacache?
28. Did Alice convince Matthew to buy the ice cream or did she buy it without him knowing?
29. Was grandpa reading the newspaper when the children got home?
30. Since grandma didn't have enough eggs, what did she decide to make, cookies or pancakes?
31. Did Matthew start from home and Alice join him along the road?
32. Did Matthew get angry with Alice or with grandma?

Repeated questions

1. Alice and Matthew aren't at grandma's house, are they?
2. Grandma didn't tell Matthew that she needed 10 eggs, did she?
3. Grandma didn't tell the children to buy an ice cream, did she?
4. Do the children have to ride along a rise before arriving in town, don't they?
5. Is it spring or summer?
6. Did the shopkeeper donate the egg to the children or did they buy it?
7. The children stopped playing in the woods before going back home, didn't they?
8. Did the children eat the ice cream after or before lunch?
9. Did grandma make the beds or wash the clothes when the children were in town?
10. Did Alice eat a strawberry ice cream?
11. Did both the children eat the ice cream or only Alice?
12. Matthew and Alice aren't cousins, are they?
13. Is grandma's house in the countryside or in the mountain?
14. Did Matthew say that Alice lost the money or that he did?
15. Is the grandma's dog name Bobby or Toby?
16. Is the ice cream parlour in town or near grandma's house?
17. Grandma didn't tell the children to ride the bike to hurry up, did she?
18. Did grandma want to make a cake with jam or a chocolate cake?
19. Matthew and Alice met their friends when they arrived in town, didn't they?
20. Alice didn't tell grandma that Matthew broke the eggs falling off the bike, did she?
21. Does grandma have a parrot or a cat at home?
22. Matthew's and Alice's mother are sisters, aren't they?
23. Didn't grandma believe either of them both or did she punish Matthew?
24. Matthew and Alice didn't eat the ice cream when they arrived in town, did they?
25. Matthew has a younger brother, doesn't he?

26. Is it going to rain or is it sunny?
27. The ice cream gives Matthew stomacache, doesn't it?
28. Did Alice buy the ice cream without Matthew knowing or did she convince him to buy it?
29. Grandpa was reading the newspaper when the children got home, wasn't he?
30. Since grandma didn't have enough eggs, what did she decide to make, pancakes or cookies?
31. Matthew didn't start from home alone and Alice join him along the road, did he?
32. Did Matthew get angry with grandma or with Alice?

2.4.2. The BISS-B22

First series of questions

1. Are Alice and Matthew at grandma's house?
2. Do the children have to ride along a rise before arriving in town?
3. Is it summer or spring?
4. Did they buy an egg or did the shopkeeper donate it to the children?
5. Did grandma wash the clothes or make the beds when the children were in town?
6. Did Alice eat a strawberry ice cream?
7. Did only Alice eat the ice cream or both the children?
8. Is grandma's house in the mountain or in the countryside?
9. Did Matthew say that he lost the money or that Alice did?
10. Is the grandma's dog name Toby or Bobby?
11. Did grandma want to make a chocolate cake or a cake with jam?
12. Did Alice tell grandma that Matthew broke the eggs falling off the bike?
13. Does grandma have a cat or a parrot at home?
14. Are Matthew's and Alice's mother sisters?

15. Did grandma punish Matthew or didn't she believe either of them?
16. Did Matthew and Alice eat the ice cream when they arrived in town?
17. Does Matthew have a younger brother?
18. Is it sunny or is it going to rain?
19. Does the ice cream give Matthew stomacache?
20. Did Alice convince Matthew to buy the ice cream or did she buy it without him knowing?
21. Was grandpa reading the newspaper when the children got home?
22. Since grandma didn't have enough eggs, what did she decide to make, cookies or pancakes?

Repeated questions

1. Alice and Matthew aren't at grandma's house, are they?
2. Do the children have to ride along a rise before arriving in town, don't they?
3. Is it spring or summer?
4. Did the shopkeeper donate the egg to the children or did they buy it?
5. Did grandma make the beds or wash the clothes when the children were in town?
6. Did Alice eat a strawberry ice cream?
7. Did both the children eat the ice cream or only Alice?
8. Is grandma's house in the countryside or in the mountain?
9. Did Matthew say that Alice lost the money or that he did?
10. Is the grandma's dog name Bobby or Toby?
11. Did grandma want to make a cake with jam or a chocolate cake?
12. Alice didn't tell grandma that Matthew broke the eggs falling off the bike, did she?
13. Does grandma have a parrot or a cat at home?
14. Matthew's and Alice's mother are sisters, aren't they?
15. Didn't grandma believe either of them both or did she punish Matthew?

16. Matthew and Alice didn't eat the ice cream when they arrived in town, did they?
17. Matthew has a younger brother, doesn't he?
18. Is it going to rain or is it sunny?
19. The ice cream gives Matthew stomacache, doesn't it?
20. Did Alice buy the ice cream without Matthew knowing or did she convince him to buy it?
21. Grandpa was reading the newspaper when the children got home, wasn't he?
22. Since grandma didn't have enough eggs, what did she decide to make, pancakes or cookies?

2.5. Measures

2.5.1. Memory measures

- *Free recall*: this index refers to the amount of information correctly reported during the immediate recall of the story. Each information unit can be scored with 1 or 0.5 point, depending on the completeness of the elements recalled. The range is 0-49.
- *Confabulation*: the number of spontaneous memory errors committed by the child during the immediate recall of the story. Memory errors may be differentiated in *Distortions* and *Fabrications*. Distortions relate to errors about details of the story (e.g. to attribute an action to the wrong character, to mistake the recipient of an action, to confuse or exaggerate some actions). Fabrications refer to completely fictional elements not present in the story and not consistent with the script or additional details that change the course of story (e.g. to attribute emotions to the characters when not mentioned in the story, insert in the narrative actions not originally described or things unsaid). One point is given to each distortion or fabrication detected.

- *Recognition memory*: this index refers to the number of correct answers to the unbiased specific questions proposed during the investigation.

2.5.2. Suggestibility measures

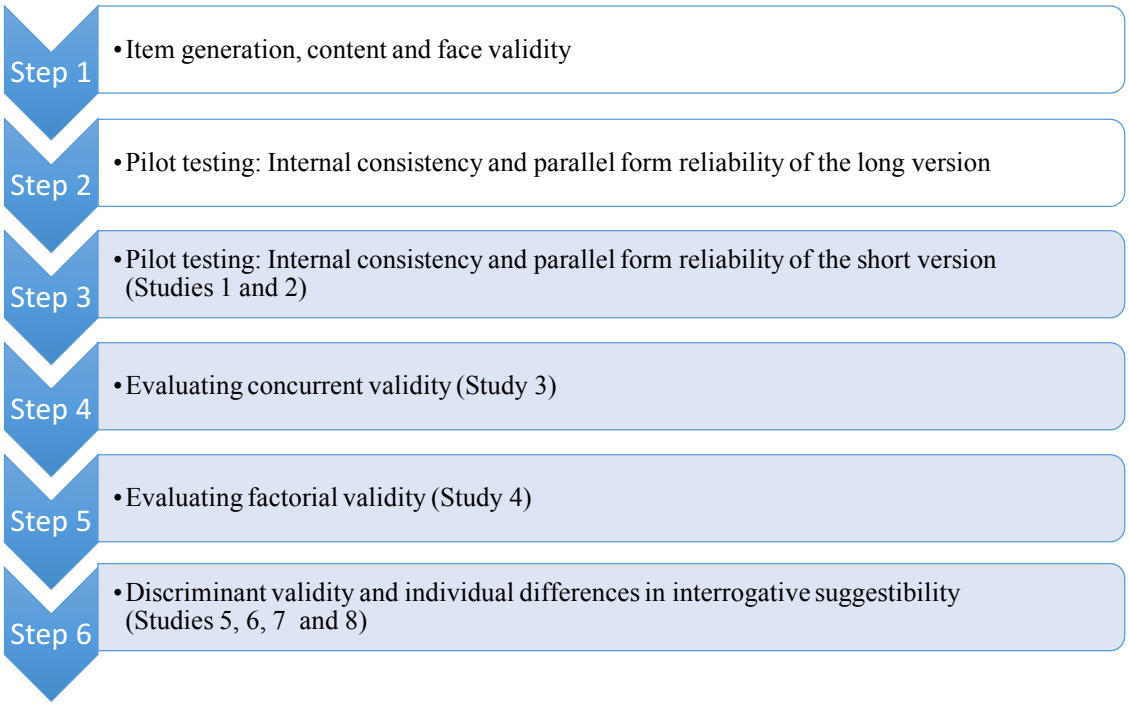
- *Yield*: the number of leading questions to which the child has succumbed during the administration of the first series of questions. The score can vary from 0 to 14 for the short form of the test (BISS A22 or B22), from 0 to 16 for the long form (BISS A32 or B32).
- *Shift*: the number of times the child has changed his or her response to the repeated questions after the negative feedback. Since all the questions are considered, both suggestive and "true" questions, this score can vary from 0 to 22 for the short form of the test (BISS A22 or B22), from 0 to 32 for the long form (BISS A32 or B32). Additionally, it is possible to discriminate three types of Shift according to the direction of the change: *Positive Shift* is considered when the child changes his or her answer from a first wrong one to the correct one after the interviewer's feedback; *Negative Shift* occurs when the child changes the first correct response with an incorrect one after the interviewer's feedback. Neutral Shift refers to the change from a wrong answer to a different answer that, however, is still wrong (e.g. in answering the question "Is the grandma's dog name Toby or Bobby?" the child says first "Toby" and then "Bobby", even though there are no dogs mentioned in the story).
- *Total Suggestibility*: this score is computed by summing the Yield and the Shift scores. It may vary from 0 to 36 in the short forms of the test (BISS A22 o B22), or from 0 to 48 in the long forms (BISS A32 o B32).
- *Post-event misinformation effect:* this index refers to the potential presence in the delayed narrative of the story of elements induced by suggestive questions. It is possible to identify the question that generate the error and to differentiate

between the four categories of questions (control Yes/no question; control alternative question; suggestive Yes/no question; suggestive alternative question). This is quite different from the misinformation effect (Loftus, 2005), since the classical procedure is experimental in nature and includes three stages: the presentation of a story, generally in an audio-visual mode, the provision of misleading information and a check on the effect by a recognition task. We introduce a change in this procedure, implementing a new step after the interview with closed questions, asking the children to provide a delayed recall of the story and coding the interference of suggested information in their free-narrative accounts.

2.6. The instrument development

The process of instrument development of the Bicocca Suggestibility Scales test-battery is shown in Figure 2.2.

Figure 2.2. The Bicocca Suggestibility Scales instrument development



The first two steps refer to studies not discussed in the present thesis. They are briefly exposed below to provide the framework of the BISS validation process.

The project started in 2010 when Claudia Caprin created a new instrument for assessing the individual vulnerability to interrogative suggestibility. She invented the stories and the questions and began the empirical validation of the tool with Italian children.

Two years later, Caprin and Benedan (2012) assessed the internal consistency of both the BISS-A32 and BISS-B32, which were administered to 396 and 453 children, respectively. The data of the children who scored below the 25th percentile in Raven's Progressive Matrix Matrices (CPM; Raven, 1996), and of those who obtained a score below the 20th percentile in the Go-No Go test (Marzocchi, Re, & Cornoldi, 2010) and in the Test of Auditory Attention (TAU; Marzocchi, Re, & Cornoldi, 2010), were excluded from the sample. The final sample comprised the BISS-A32 data of 322 children (150 males), and the BISS-B data of 343 children (161 males), all aged between 6 and 11 years.

Cronbach's alphas were computed for the memory and suggestibility scores of BISS-A32 and BISS-B32, and are shown in Table 2.1. According to Kline's (2013) and Nunnally's (1978) recommendations, and considering that this validation study concerns the early stages of research, coefficients of .70 or higher have been considered as indicators of good enough reliability of the instrument.

The results suggested that both the BISS forms have a good internal consistency for the Yield and the total suggestibility scales; internal consistency was also acceptable for the Shift scale, while the reliability of recognition memory measure was not acceptable at all, being the Cronbach's alpha values considerably lower than the aforementioned cut-off criteria of .70.

Then, a shorter version of the instrument was created. The results of the internal consistency analysis led to the choice of the items to delete. The theoretical internal consistency was then assessed by considering only the items that would be kept in the

shorter form. The new alpha values were higher than the previous ones for Yield, Shift and total suggestibility, while the Cronbach's alpha of the recognition memory measure got worse.

Table 2.1. Internal consistency of the BISS scales.

	N. of items	α BISS-A	α BISS-B
BISS-32 items:			
Recognition memory	16	.46	.51
Yield	16	.89	.87
Shift	32	.73	.76
Total suggestibility	48	.87	.87
BISS-22 items:			
Recognition memory	8	.41	.33
Yield	14	.89	.88
Shift	22	.76	.77
Total suggestibility	36	.88	.88

Items deleted:
- BISS-A32: 1, 2, 3, 4, 6, 7, 8, 16, 17, 31
- BISS-B32: 2, 3, 7, 8, 12,16, 17, 19, 31,32

Subsequently, a study was conducted to assess the reliability of the parallel forms for the long version of the BISS. One-hundred-and-eighty children aged 7 to 11 years participated in this study. Pearson's bivariate correlations were performed. The results suggested that the reliability for parallel forms was satisfactory for both the memory measures (free recall: r = .48, p <.001; recognition memory: r = .33, p <.001), and the three suggestibility scales (Yield: r =.56, p <.001; Shift: r =.54, p <.001; total suggestibility: r =.65, p <.001).

Given the results of these preliminary studies, the BISS appears to reliably measure different aspects of children's suggestibility. Therefore, it is not unlikely that the scale could become a useful measure of individual children's vulnerability to suggestion after having ascertained its validity.

The following steps of the BISS validating process are the content of the empirical studies presented in this thesis in the following chapters.

CHAPTER 3

The psychometric properties of the Bicocca Suggestibility Scales

The reliability of children's testimonies is a serious concern. Although it is generally recognised that situational factors may contribute to the increased vulnerability to suggestion, it is contentious that suggestibility is a dispositional factor which makes some people more vulnerable than others to the effect of leading questions (Baxter, 1990). The high variability in children's vulnerability to suggestibility observed within and across studies showed that, beyond the characteristics of the interviewer behaviour and the interview context, some children are more likely than others to be misled by leading questions. Moreover, the literature on consistency of children's memory and suggestibility has revealed that children behave in predictable ways. As already shown in previous works with standardised scales of memory and suggestibility (BTSS, VSSC, GSS; see chapter 2 for a detailed description), it is possible to predict children's ability to accurately report an event and their susceptibility to misleading information provided by interviewers.

The idea underlying the construction of a standardised memory and suggestibility scale is that the processes and outcomes observed in an eyewitness interview about one specific event may be considered as stable indicators of a unique and global trait-like construct (Endres, 1997). This would affect children's performance across different eyewitness tasks and different events related to the same underlying trait. Moreover, individual differences in the stability of performance across different recall tasks and different events would add information in the extent to which the instrument measures the construct.

So that an instrument be considered reliable, its scores should be interpreted consistently across different situations. When the same instrument is administered twice to the same people, it should produce similar scores at both points in time (test-retest reliability) (Field, 2009).

Considering the nature of the subject matter of interest, namely memory and suggestibility, it is believed that the administration of the same stimulus twice to the same subject would not be optimal. This is because people usually retain some memory of the content of the story and this could affect their subsequent performance. However, it is advantageous for both clinical and research purposes being able to assess memory and suggestibility on more than one occasion (Gudjonsson, 1997). Therefore, two parallel versions of the BISS have been created.

The concept of stability is of interest because it may inform about the predictability of a certain behaviour (Wohlwill, 1973). Stability may be defined as the preservation of individual differences from one measure of a variable to one taken subsequently. In this view, unstable is equal to unpredictable (Wohlwill, 1973). Therefore, understanding whether children's performance in a memory and suggestibility task is stable may shed light on the issue of the predictability of children's vulnerability to suggestive influences. In the studies presented in this thesis, it was of interest to distinguish group stability and individual stability (Asendorpf, 1992). Group stability refers to the stability of the dimension over time (Roebers & Schneider, 2002). It is typically assessed by computing the test-retest correlation coefficient. Group stability provides also an indicator of the validity of the construct of interest (Wohlwill, 1973). Unlike group stability, individual stability focuses on variability between different individuals inside the group. In other words, it refers to the relative stability of the position of each member within that group at different times (Wohlwill, 1973). Individual stability is measured by observing the fluctuation in individuals' deviation from the mean: the individual with the highest score at Time 1 remains the highest at Time 2.

For example, the individual's performance of the person with the highest score at Time 1 is stable if this person reaches the highest score at Time 2. If many individuals' relative positions respect to the others inside the group vary largely, it signifies that the measure is unstable at the individual level and, therefore, the future performance would be less predictable. Bayley (1949) defined "lability score" the intraindividual change observed in longitudinal studies. This term would refer to the child progress in relation to his or her previous performance and could be computed by considering the standard deviation from the mean of his or her standard scores obtained across all testing points. A higher score means greater variability from the child's own central tendency. Similarly, Asendorpf (1989, 1992) introduced the concept of individual or differential stability, that may be computed with the following formula:

$$i_{12} = 1 - \frac{(z_1 - z_2)^2}{2}$$

In this equation, z_1 and z_2 represent the z-transformed scores at Time 1 and Time 2, respectively. Therefore, the individual stability score refers to the intra-individual variability of the child between two z-transformed assessments subtracted from 1. The mean of these scores would be identical to the correlation between the two assessments in the sample of interest, while the standard deviation from the mean would reflect whether all individuals of the sample are equally predictable.

In Study 1 and 2 both group stability and individual differences in stability of memory and suggestibility scores were assessed. The consistency of performance across the different BISS forms tasks was evaluated. In study 1 the reliability of the parallel forms was the main focus of interest, while Study 2 was aimed to observe the developmental changes of memory and suggestibility over a 2-year period.

3.1. STUDY 1. The reliability of the short forms

Study 1 aimed to assess the reliability of the parallel forms of the Bicocca Suggestibility Scales - short version. To investigate whether the Bicocca Suggestibility Scales is an instrument suitable to predict performance across different eyewitness tasks, both the form A and the form B of the instrument were used. The two forms were administered two weeks apart from each other by counterbalancing the presentation order.

To assess the reliability of the BISS, first the internal consistency of both the BISS-A22 and the BISS-B22 was evaluated, as well as inter-task consistency. Second, group and individual stability of memory and suggestibility scores were examined. Third, the reliability of the parallel forms of the Bicocca Suggestibility Scales was evaluated.

3.1.1. Method

Participants

One hundred and thirty-eight 6- to 11-year-old children were recruited through letters distributed to four primary schools in the Northern Italy[11]. All children had parental consent to participate, and none was identified by the teachers as having significant problems communicating in the Italian language. From the original sample, the data of 9 children were excluded because they scored below the 25th percentile in the Raven's Coloured Progressive Matrices (RCPM, Raven, 1998). The final sample comprises 129 children (66 males), (M_{age} = 8.25, SD = 1.42).

Procedure and instruments

The two parallel versions of the Bicocca Suggestibility Scales battery were used: BISS-A22 and BISS-B22. To avoid a possible bias in the presentation order, the two

[11] Alessandra Telò, Irene Falceri, Davide Giuseppe Sabadini, and Chiara Fortis collected the data.

versions of the BISS were administered in a counterbalanced way, i.e. half of the children received the BISS-A22 during the first meeting and the BISS-B22 during the second meeting; the remaining half of the children completed the BISS-B22 first and the BISS-A22 two weeks later.

For both the instruments, the following measures were computed:

– *Free recall* = the amount of information correctly reported during the immediate recall of the story (maximum = 49).

– *Recognition memory* = the number of correct answers to those specific yes/no and forced-choice questions referring to elements presented in the story. This index was computed by considering only the answers to the first series of questions (maximum = 8).

– *Yield* = the number of leading questions to which the child has given the wrong answer during the administration of the first series of questions (maximum = 14).

– *Shift* = the number of questions to which the child has changed his or her response after receiving the negative feedback (maximum = 22). According to the direction of the change, it is possible to distinguish three different kind of Shift:

- *Negative shift:* when the child firstly chose the correct answer, but changed his or her answer after the interviewer's negative feedback by selecting one wrong option (e.g. when asked "Was mum cooking or ironing?"[12] the child says "I don't know" the first time and then "cooking" or "ironing").
- *Positive shift:* when the child provided a wrong answer during the first administration of the questioning protocol and corrected himself or herself after the interviewer's negative feedback (e.g. the child says "cooking" or "ironing" the first time and then he or she says "I don't know").

[12] This is a suggestive question, therefore the question refers to elements that have not been presented in the story. Indeed, there is no reference to what the mum was doing, and the child cannot know the answer to the question "Was mum cooking or ironing?" Hence, the correct answer to this question is "I don't know".

- *Neutral shift*: when the child provided a wrong answer during the first administration of the questioning protocol and changed his or her answer after the interviewer's negative feedback, by selecting the other wrong alternative suggested by the interviewer (e.g. the child once says "cooking" and once says "ironing").

– *Total suggestibility* = this index provides an overall estimate of the child's vulnerability to interrogative suggestibility. It is obtained by summing the Yield and the Shift score (maximum = 36).

– *Misinformation effect* = the number of the details suggested during the questioning phase that are subsequently reported in the delayed free recall of the story.

3.1.2. Data analysis

The statistical analyses of the present study were performed using RStudio software, version 1.1.383. The *psych* package was used (Revelle, 2017, 2017). The internal consistency (reliability) of the suggestibility scales was measured through ordinal alpha coefficients because the observed variables were scored dichotomously (see Gadermann, Guhn, & Zumbo, 2012; Zumbo, Gadermann, & Zeisser, 2007). Ordinal alpha provides a suitable estimate of the theoretical reliability, regardless of the magnitude of the theoretical reliability, the number of scale points, and the skewness of the scale point distributions. Therefore, ordinal alphas were computed to estimate reliability more accurately through an unbiased estimator of the theoretical reliability for dichotomous data (Gadermann et al., 2012). Coefficients of .70 or above were considered satisfactory (Kline, 2013).

3.1.3. Results

First, the children's performance on both the BISS-A22 and BISS-B22 memory and suggestibility scales was considered. Table 3.1 displays the descriptive information about all the indexes considered, along with reliability scores. By considering the recognition memory scores of both the BISS-A22 and the BISS-B22, there is a considerable ceiling effect and lack of variability in the children's performance when answering to the "true" questions.

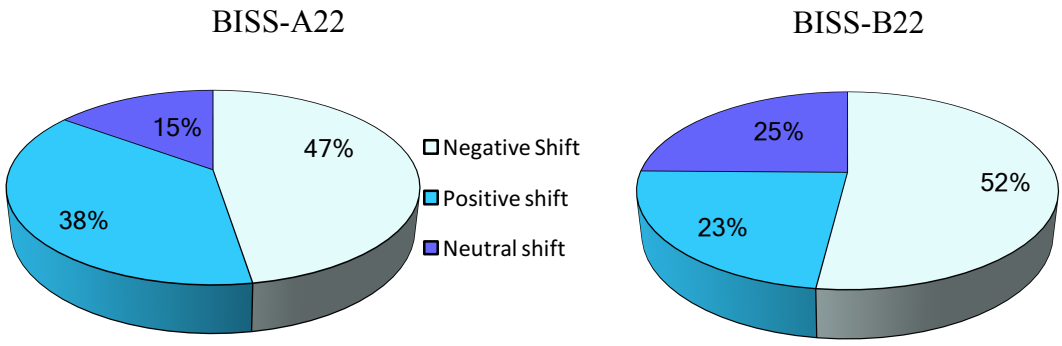
Table 3.1. Descriptive and reliability analyses on the BISS-A22 and BISS-B22

	Range	M	SD	Skewness	Kurtosis	Ordinal α
BISS-A22						
Free recall	2.5 - 45.0	23.79	9.77	-0.11	-0.73	NA
Recognition memory	3.0 - 8.0	6.88	1.03	-0.79	0.40	.40
Yield	0.0 - 14.0	6.20	4.71	0.02	-1.49	.97
Shift	0.0 - 16.0	6.04	3.62	0.20	-0.47	.83
Total suggestibility	0.0 - 27.0	12.24	7.69	-0.13	-1.29	.94
BISS-B22						
Free recall	0.0 - 40.5	22.88	9.06	-0.25	-0.68	NA
Recognition memory	3.0 - 8.0	6.92	1.18	-1.08	0.75	.64
Yield	0.0 - 13.0	5.80	4.18	0.00	-1.49	.95
Shift	0.0 - 17.0	5.03	3.95	0.89	0.27	.89
Total suggestibility	0.0 - 28.0	10.83	7.02	0.07	-1.12	.93

Ordinal alpha values were found to be satisfactory for the three interrogative suggestibility scores, indicating good internal consistency for both the BISS-A22 and BISS-B22. Conversely, the recognition memory measure was not reliable enough. Even though the internal consistency of the recognition memory score for the BISS-B22 (ordinal $\alpha = .64$) was better than those of the BISS-A22 (ordinal $\alpha = .40$), these values did not reach the at least acceptable cut-off point of .70.

Figure 3.1 displays the percentage of each kind of shift on the total Shift score. It can be noted that about half of the changes made are in the negative direction. About once every two changes made by the child, the interpersonal pressure of the interviewer's negative feedback led the child to modify the correct answer with the wrong one. This is consistent between the two forms of the BISS.

Figure 3.1. Percentages of the different kinds of Shift made by the children



Then, a series of Pearson's bivariate correlations were performed between the memory and suggestibility measures assessed with BISS-A22 and the same measures assessed with BISS-B22. Intra-task and inter-task correlation coefficients are displayed in Table 3.2.

The correlation coefficients indicating parallel forms reliability, or group stability, were statistically significant. The Yield score had a good reliability (r = .83, p <.001), and even free recall (r = .76, p <.001) and total suggestibility (r = .78, p <.001) reliability were acceptable. Although the Shift score correlation coefficient (r = .61, p < .001) was significant, the reliability was questionable because it did not reach the standard cut-off point of .70. The recognition memory score seemed to be an unacceptably reliable index. Then, individual and differential stability[13] scores were calculated to assess the stability

[13] The Asendorpf's (1989; 1992) formula was used: $i_{12} = 1 - \frac{(z_1 - z_2)^2}{2}$

of individual's relative position within the group between the two assessments. First, the BISS scores were transformed into z scores. Free recall individual stability was .77 (*SD* = .32), while recognition memory score individual stability was .26 (*SD* = 1.16). Hence, free recall measure is stable enough when measured with the two different forms of the BISS. Conversely, the recognition memory score presented a quite heterogeneous distribution of individual stability.

Regarding the suggestibility measures, individual stability was .82 (*SD* = .36) for Yield score, .63 (*SD* = .57) for Shift score, and .79 (*SD* = .35) for Total Suggestibility score.

Table 3.2. Intra-task and inter-task Pearson's correlations among the subscales of the BISS-A22 and BISS-B22

		F. R. [B22]	R. M. [B22]	Yield [B22]	Shift [B22]	T. S. [B22]
Inter-task						
F. R.	[A22]	.76***	.24**	-.50***	-.46***	-.56***
R. M.	[A22]	.26**	.22*	-.08	-.03	-.06
Yield	[A22]	-.43***	-.08	.83***	.44***	.73***
Shift	[A22]	-.47***	-.06	.62***	.61***	.71***
T. S.	[A22]	-.49***	-.08	.80***	.55***	.78***
Intra-task						
F. R.		1.00	.26**	-.50***	-.32***	-.48***
R. M.		.28**	1.00	-.02	-.12	-.08
Yield		-.49***	-.01	1.00	.50***	.87***
Shift		-.56**	-.16	.70***	1.00	.86***
T. S.		-.56***	-.08	.94***	.90***	1.00

In the intra-task correlations section, the coefficients above the diagonal refer to BISS-B22, while the coefficients below the diagonal refer to BISS-A22.

F. R. = free recall; R. M. = recognition memory; T. S. = total suggestibility.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Finally, the misinformation effect was considered. Table 3.3 displays the frequency of which the children showed the phenomenon for both the BISS-A22 and the BISS-B22. Most of the children did not present the misinformation effect, i.e. did not

include any information suggested by the interviewer during the questioning phase in their delayed recall of the story. Specifically, 71% and 72% of the children did not show the misinformation effect when recalling the BISS-A22 and the BISS-B22, respectively. However, it should be noted that many children who showed the misinformation effect for the BISS-A22 were not the same who did it for the BISS-B22. Indeed, only 54% of the children did not show the misinformation effect at all. About 35% of the children in the sample showed the phenomenon in only one version of the test, while 11% of them displayed the effect for both the stories.

Table 3.3. Frequency of the misinformation effect for both the BISS-A22 and the BISS-B22.

		BISS-B22					
		0	*1*	*2*	*3*	*4*	*Total*
BISS-A22	*0*	69	21	2	0	0	92
	1	19	4	3	1	0	27
	2	3	1	0	1	1	6
	3	0	2	1	0	0	3
	Total	91	28	6	2	1	128

3.2. STUDY 2. Children's vulnerability to suggestibility: stability and consistency over time

Study 1 showed the reliability of the parallel forms of Bicocca Suggestibility Scales – short version. In Study 2 the aim was to evaluate children's memory and suggestibility performance over time. According to the findings of the previous study, it is reasonable that interrogative suggestibility is based on some individual's stable tendencies to respond in a certain way to a given situation. It is therefore of interest to examine the stability over time of children's individual differences in memory and suggestibility.

Moreover, long-term stability may help in understanding how much a particular trait changes over a given time period (Asendorpf, 1992). The main limitation of cross-sectional data is the lack of direct information about the individual's change and the constancy of an individual's standing relative to some referent group across age. To investigate the changes in behaviour occurring along the dimension of age it is necessary to consider the differences in the performance of the same individual at two different times. To overcome this limitation and to examine developmental changes in memory and suggestibility, we assessed the same group of children at two different times over a two-year period.

The present study had three main aims. The first referred to the evaluation of changes in memory and suggestibility over time within the same group of children. The second aim was the assessment of the stability of individual differences between children after a two-year delay. The third aim was related to providing further support about the validity of the parallel forms of the Bicocca Suggestibility Scales.

3.2.1. Method

Participants

Participants included 28 children (16 males, 12 females) who remained from the original sample of 30 children enrolled in a primary school of Northern Italy[14]. They were tested at age 8 (M_{age} = 8.59, *SD* = .36) and age 10 (M_{age} = 10.68, *SD* = .31).

Procedure and instruments

The two parallel versions of the Bicocca Suggestibility Scales Battery were used. BISS-A22 was administered at time 1, while BISS-B22 was administered two years later. For both the instruments, the following measures were computed:

- *Free recall* = the amount of information correctly reported during the immediate recall of the story (maximum = 49).
- *Recognition memory* = the number of correct answers to the true specific yes/no and forced-choice questions during the administration of the first series of questions (maximum = 8).
- *Yield* = the number of leading questions to which the child has succumbed during the administration of the first series of questions (maximum = 14).
- *Shift* = the number of questions to which the child has changed his or her response after the negative feedback (maximum = 22).
- *Total suggestibility* = the sum of Yield and Shift score (maximum = 36).
- *Misinformation effect* = the number of details suggested during the questioning phase and subsequently reported in the delayed free recall of the story.

[14] Chiara Spinelli collected the data on both occasions.

3.2.2. Results

First, we considered children's performance on both the BISS-A22 and BISS-B22 memory and suggestibility scales. Table 3.4 displays the descriptive information about all the measures considered. As previously noted in Study 1, the recognition memory scores showed a ceiling effect for both the BISS-A22 and the BISS-B22.

Table 3.4. Descriptive information for memory and suggestibility measures

	Range			M	SD	Skewness	Kurtosis
BISS-A22							
Free recall	5.50	-	35.00	21.14	8.14	-0.46	-0.41
Recognition memory	5.00	-	8.00	7.18	0.98	-1.14	0.46
Yield	0.00	-	14.00	8.93	5.07	-0.56	-1.32
Shift	2.00	-	12.00	6.25	3.01	0.25	-1.18
Total suggestibility	2.00		25.00	15.18	7.35	-0.45	-1.2
Misinformation effect	0.00	-	4.00	0.64	0.87	2.25	7.38
BISS-B22							
Free recall	16.50	-	32.00	26.34	3.68	-0.63	0.75
Recognition memory	7.00	-	8.00	7.93	0.26	-3.52	11.18
Yield	0.00	-	9.00	3.71	2.37	0.06	-0.84
Shift	0.00	-	6.00	2.39	1.66	0.77	-0.38
Total suggestibility	1.00	-	15.00	6.11	3.57	0.55	-0.27
Misinformation effect	0.00	-	2.00	0.46	0.58	0.77	-0.34

Multiple t-tests were used to consider differences in gender for memory and suggestibility scores (*p*s >.05). Subsequent analyses were therefore conducted on the collapsed data.

Then the relationship between age, memory and suggestibility was investigated within groups for each measurement point. Pearson's bivariate correlations showed that age was not related to memory performance at T1 (Free recall: r = .07, p = .73; Recognition memory: r = .21, p = .28) and T2 (Free recall: r = -.19, p = .32; Recognition

memory: r = .07, p = .72), nor to interrogative suggestibility measures at T1 (Yield: r = .06, p = .75; Shift: r = .28, p =.15; Total suggestibility: r = .16, p =.42) and T2 (Yield: r = .06, p = .75; Shift: r = .19, p = .33; Total suggestibility: r = .08, p =.70).

The misinformation effect has been considered separately and the frequency of the occurrence of this phenomenon is shown in Table 3.5. Then, the information about occurrence of the misinformation effect has been dichotomized as presence or absence. The Mc Nemar's X^2 was performed and no significant difference was found (p = .79) between the children who showed the misinformation effect at time 1 (50%) and those who showed it at time 2 (43%). It was then examined if the children who introduced suggested information on their subsequent recall at time 1 were the same who incurred in this phenomenon at time 2. Only 29% of the sample did not show the phenomenon in both occasions, while the others changed their behaviour either in a positive or in a negative way.

Table 3.5. Frequency of the misinformation effect for both the BISS-A22 and the BISS-B22.

		BISS-B22					
		0	*1*	*2*	*3*	*4*	*Total*
BISS-A22	*0*	8	5	1	0	0	14
	1	6	6	0	0	0	12
	2	1	0	0	0	0	1
	3	0	0	0	0	0	0
	4	1	0	0	0	0	1
	Total	16	11	1	0	0	28

Then age was considered as within-subject factor, and a series of paired samples t-tests were performed to assess the effect of the individuals' development underlying the passing of time on memory and suggestibility scores. After a two-year delay, there was a statistically significant improvement in children's memory performance, both in terms of the amount of accurate information spontaneously recalled (t (27) = -3.95, p = .001, d = -

0.75) and in the accuracy when answering specific questions about elements originally present in the story (t (27) = -4.28, p < .001, d = -0.81). Regarding interrogative suggestibility, there was a statistically significant decrease in the likelihood to succumb to misleading questions (t (27) = 8.41, p < .001, d = 1.59) and to change answers after the adult's negative feedback (t (27) = 8.62, p < .001, d = 1.63), as well as in the overall suggestibility level (t (27) = 9.88, p < .001, d = 1.87). The probability of getting a significant result due to chance was considered and the Bonferroni correction was applied for multiple tests. The critical p-value was p = .01 and, therefore, all the results were still significant after considering this adjustment.

After having assessed children's responses over time, their individual performance and the stability over time of their memory ability and vulnerability to suggestion were evaluated. Pearson's bivariate correlations were conducted between the memory and suggestibility measures assessed with BISS-A22 and the same measures assessed with BISS-B22 after a two-year delay. Free recall scores were significantly correlated (r = .52, p < .001), while the positive relationship between accuracy scores in answering specific questions did not reach the statistical significance (r = .34, p = .08). Yield scores (r = .86, p < .001), as well as Shift scores (r = .62, p < .001) and total suggestibility scores (r = .82, p < .001) were strongly related. The Yield scale showed a good stability coefficient, while Free Recall and Shift scores could be considered modestly stable over time.

Figure 3.2 displays the distribution of free recall scores for each time at the two assessment times. It is evident that, even though most of the children's performance is better at time 2 than at time 1, some children recalled a fewer elements on the second assessment.

The distribution of the recognition memory scores is shown in Figure 3.3. Almost all the children reached the 100% of accuracy in answering specific closed-questions related to elements present in the story.

Figure 3.2. Free recall scores of each child at time 1 and 2

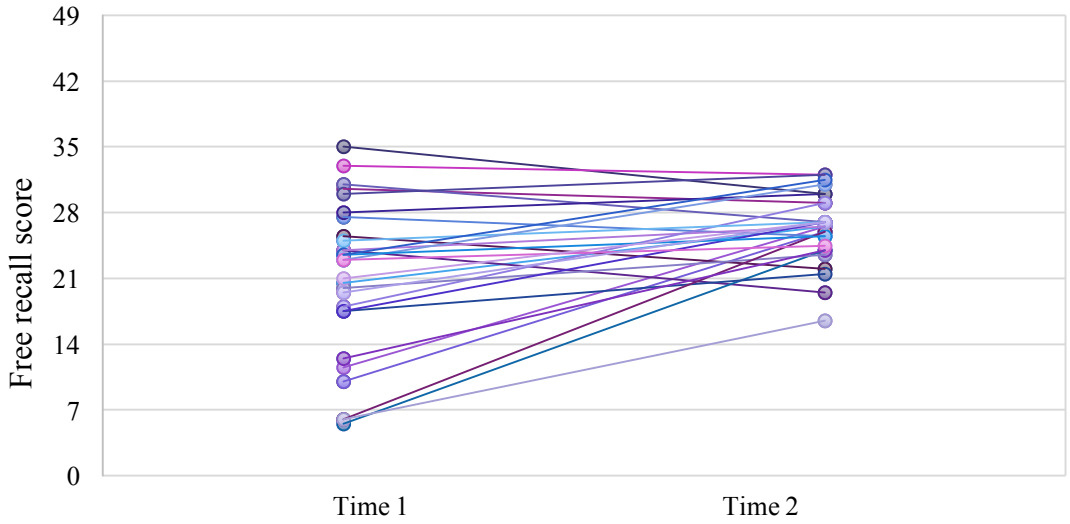
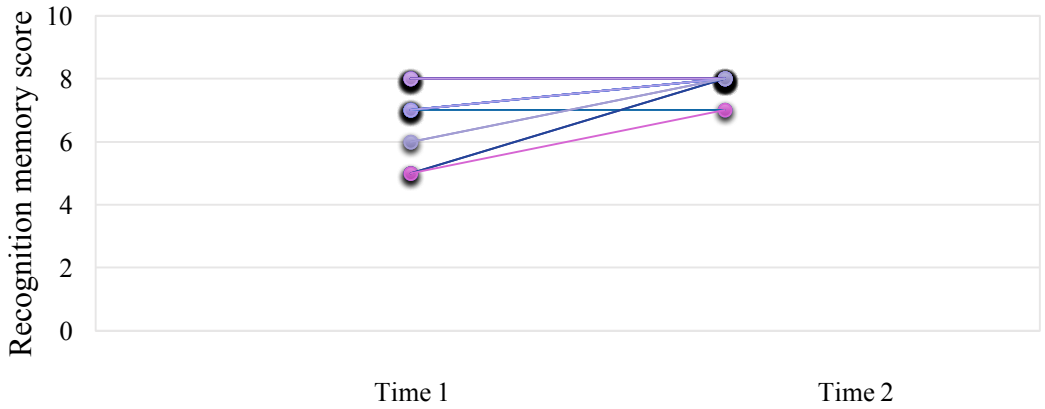


Figure 3.3. Recognition memory scores of each child at time 1 and 2



Finally, to examine whether individual's relative position within the group remained stable between the two assessment times, individual and differential stability scores were calculated. The BISS scores were transformed into z scores. Individual stability for free recall was .54 (*SD* = .57), while individual stability of the recognition memory score was .36 (*SD* = 1.26). These values indicate a quite heterogeneous distribution of individual stability of memory measures. Individual stability for Yield score was .86 (*SD* = .26), while individual stability of Shift score and Total Suggestibility

score were .64 (*SD* = .74) and .83 (*SD* = .25), respectively. The distribution of the Yield, Shift and Total Suggestibility scores are shown in Figure 3.4, 3.5 and 3.6, respectively.

Figure 3.4. Yield scores of each child at time 1 and 2

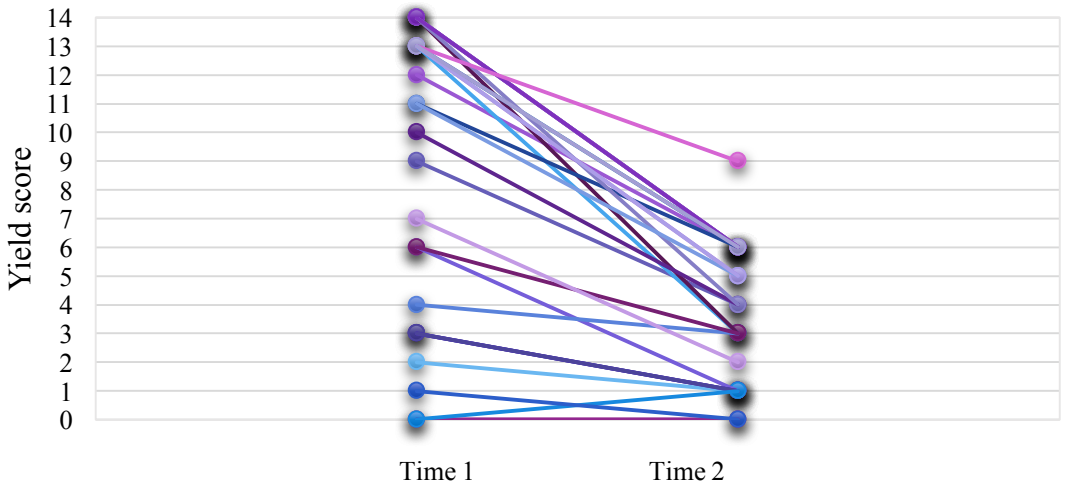


Figure 3.5. Shift scores of each child at time 1 and 2

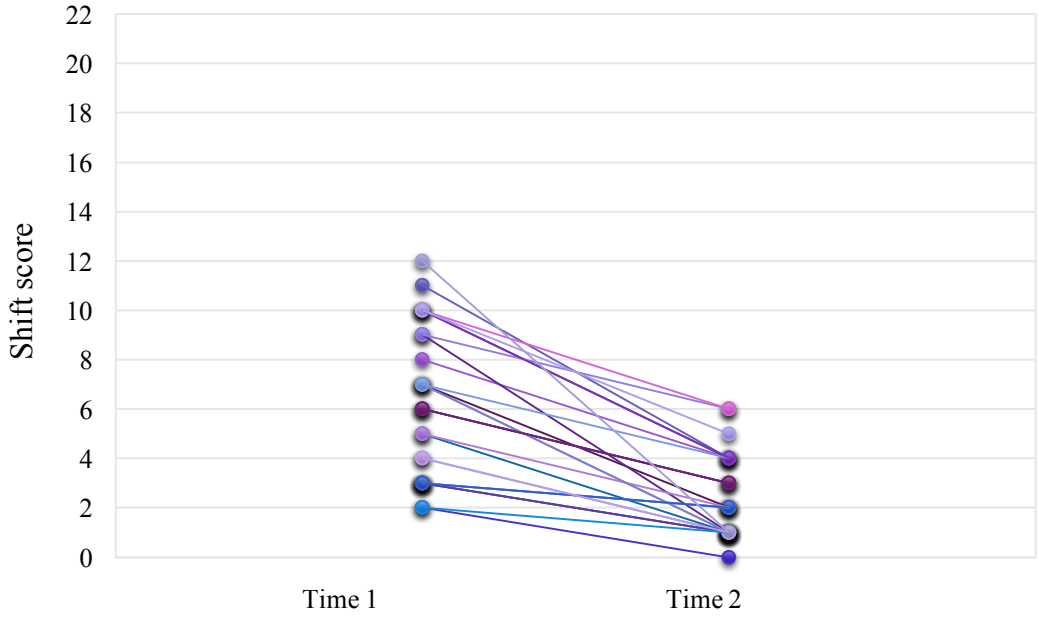
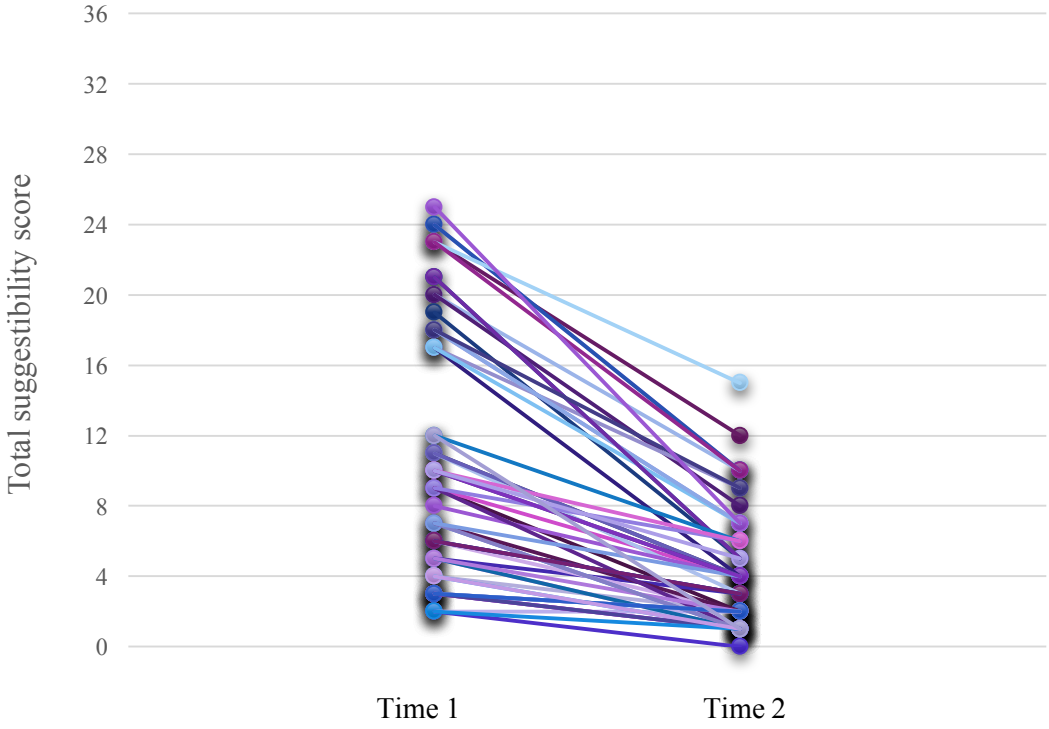


Figure 3.6. Total suggestibility scores of each child at time 1 and 2.



3.2.3. Discussion of Studies 1 and 2

In Study 1 and Study 2 the test-retest reliability of the memory and suggestibility scores between the BISS-A22 and the BISS-B22 was assessed. Therefore, it was of interest to ascertain whether the items that were supposed to measure the same general construct were adequate in producing similar scores. This aim was achieved for the Yield score, which presented satisfactorily test-retest reliability and a good internal consistency for both the forms of the BISS, with ordinal α ranging between .95 and .97. Moreover, this values are higher than those of the Italian version of the Bonn Test of Statement Suggestibility (BTSS), which reported a Kuder-Richardson coefficient of .78 (Rossi et al., 2011). Similarly, the BISS internal consistency for the Yield score was higher than those of the Gudjonsson Suggestibility Scales (GSS), whose values ranged between .66 and .72 for the Italian version (Bianco & Curci, 2015).

Although the correlations between the BISS-A22 and the BISS-B22 Shift factor were high and significant, the test-retest reliability was not optimal, but could be considered acceptable considering the low variability of these scores. Moreover, this result is better than the test-retest coefficient of the original version of the BTSS, which was .32 (Endres, 1997). The internal consistency of the Shift factor may be considered good enough, with ordinal α ranging between .83 and .89. Again, these values are higher than those of the Italian version of the BTSS, with a Kuder-Richardson coefficient of .77 (Rossi et al., 2011), and those of the Italian version of the GSS, with Cronbach's alpha values ranging between 53 and .58 (Bianco & Curci, 2015).

Conversely to the results of the suggestive measures, the recognition memory score of both the BISS-A22 and the BISS-B22 did not show satisfactory test-retest reliability nor adequate internal consistency. It should be noted that this measure was affected by a considerable ceiling effect. This lack of variability in the children's performance when answering to the "true" questions was more evident in Study 2 for the BISS-B22, which was administered two years later than the BISS-A22. Therefore, at time 2, the children showed a high level of accuracy in answering specific-closed questions about elements actually present in the story read. This lack of variability could explain the low ordinal alpha values of the recognition memory scores.

Recognition memory and free recall scores were significantly related when evaluating both inter-task and intra-task correlation coefficients. This means that an underlying general memory ability rendered account for the children's performance in free recall and recognition memory tasks. The same happened for the three interrogative suggestibility scores. In particular, Yield and Shift scores were highly associated. In other words, the higher the tendency to give a wrong answer to a misleading question, the higher the odds to succumb to the interpersonal pressure to change the previous answer, regardless of whether it was correct or not. This association was observed when considering the same scale, but even when crossing the measures of the two forms of the

BISS. Additionally, there was a negative association between free recall and the three interrogative suggestibility scores intra-task and inter-task. Hence, a better memory for the story would decrease the probability to be misled by suggestive questions and to shift the first answer to a question when interviewed about the same story. This finding is in line with existing literature (Danielsdottir et al., 1993; Finnilä et al., 2003; McFarlane et al., 2002; Scullin & Bonner, 2006). It seems plausible that a robust memory trace for a given even would reduce uncertainty, and therefore this would decrease the vulnerability to interrogative suggestibility (Gudjonsson & Clark, 1986). However, the inter-task correlations results between free recall and interrogative suggestibility bring evidence to the hypothesis that free recall is a good indicator of the general memory ability of the child. They would also suggest that the general memory ability is a protective factor against the vulnerability to interrogative suggestibility. This finding would provide some support to the conclusion that interrogative suggestibility is a unique trait that manifests when answering to misleading questions and in responding to repeated questions.

Moreover, Study 2 investigated stability and consistency of children's event recall and their vulnerability to interrogative suggestibility for two different stories after a two-year delay. Consistent with cross-sectional findings on children's development of memory ability and suggestibility proneness (Danielsdottir et al., 1993; Endres, 1997), data provided support to the hypothesis that with development there is a noticeable improvement in memory recall along with a decrease in vulnerability to both misleading questions and adult's negative feedback.

Regarding the long-term group stability of the memory scores, the results of Study 2 confirmed the significant high test-retest correlations shown by previous studies with younger children (Lehman et al., 2010; Roebers & Schneider, 2002; W. Schneider & Sodian, 1991; W. Schneider & Weinert, 1995). Moreover, a strong group stability as well as individual stability was found for children's tendency to yield to misleading questions, supporting the previous finding of Lehman et al. (2010). Finally, children's likelihood of

changing their responses after an adult's negative feedback was stable over a two-year interval. The individual stability measure was affected by the floor effect detected at time 2, when ten-year-old children rarely shifted answer after the adult's negative feedback.

Interestingly, no significant difference was found in the tendency to exhibit the misinformation effect between time 1 and time 2. Indeed, 43% of the children introduced at least a piece of suggested information at time 2.

These findings are consistent with the hypothesis that the suggestibility is a stable trait-like characteristic. They emphasize the relevance of identifying those children with a higher vulnerability to interrogative suggestibility, because this information would allow the police or jurors to estimate the potential effects of suggestions prior to an interview.

3.3. STUDY 3. The concurrent validity

The aim of Study 3 was to assess the concurrent validity of the Bicocca Suggestibility Scales by comparing this test with the Bonn Test of Statement Suggestibility (Endres, 1997; Rossi et al., 2011). First, Pearson's correlations were computed to assess the relationship between memory and suggestibility scores measured with the two instruments. Then, to assess the agreement between two quantitative methods of measurement, the Bland and Altman's method (Bland & Altman, 1983, 1999, 2010) was applied. Specifically, Bland-Altman plots are a well-established method to check the agreement of different measurement methods, by constructing limits of agreement. This method quantifies the bias and a range of agreement, within which 95% of the differences between one measurement and the other are included (Giavarina, 2015).

3.3.1. Method

Participants

Participants included 142 children (77 males, 65 females) enrolled in 4 primary schools of Northern Italy[15]. They were aged between 6 and 10 year (M_{age} = 7.98, *SD* = 1.16).

Instruments

The BISS-A32 was used. The following measures were considered:

- Free recall: the amount of information correctly recalled (maximum = 49).
- Recognition memory: the number of correct responses to the unbiased yes/no and forced-choice questions during the administration of the first series of questions (maximum = 16).

[15] Greta Lonati, Elena Casula, Flavia Achermann, and Debora Smecca collected the data.

– Yield: the number of misleading questions to which the child has given the wrong answer during the administration of the first series of questions (maximum = 16).

– Shift: the number of questions to which the child has changed his or her response after the interviewer's negative feedback (maximum = 32).

Bonn Test of Statement Suggestibility[16]

The Bonn Test of Statement Suggestibility (BTSS; Endres, 1997; italian version: Rossi, Benatti, Pesce, Oppo, & Avato, 2011) has been used to assess the concurrent validity of the BISS.

The following memory and suggestibility measures were computed:

– Free recall: the amount of information correctly recalled (maximum = 62)[17].

– Recognition memory: the number of correct responses to the "distractor" questions (maximum = 6). In the original version of the test, distractor questions are just filler items used to disguise the real purpose of the test. These questions are not usually considered since they are leading questions in which the correct answer is suggested (e.g., "The boy's name was Oliver, wasn't it?"). However, in the present study they are used for the comparison with the recognition memory scale of the BISS-A32.

– Yield: the number of misleading questions to which the child has given the wrong answer during the administration of the first series of questions (maximum = 17). This score includes both the answers to the misleading yes/no

[16] See paragraph 1.4.2. for a detailed description.

[17] The segmentation in informative units of the BTSS has been adapted in the present study to grant a more appropriate comparison with the segmentation of the BISS-A32.

questions and to the wrongly disjunctive alternative questions. The first category refers to those questions which state an incorrect fact by suggesting that an affirmative answer is expected, while the correct answer is "no". The second category refers to those questions which present two equally non-correct answer options. Since the information required is not available to the child or it is not included between the two options, the correct answer is "I don't know".

– Shift: the number of questions to which the child has changed his or her response, unless the Shift is from "I don't know" to "no" or vice versa (maximum = 8). Eight yes/no or alternative questions are immediately repeated to the child, irrespective of the answer just given (e.g., "Are you sure? Did he want to buy apples or bread?"). Even if no negative feedback was provided, these questions convey the message that the answer the subject has just given is not accepted and that he or she is expected to change this answer.

Procedure

To avoid a possible bias in the presentation order, the BISS-A32 and BTSS were administered in a counterbalanced way, i.e. half of the children received the BISS-A32 during the first meeting and the BTSS during the second meeting; the remaining half of the children completed the BTSS first and the BISS-A32 two weeks later.

3.3.2. Results

The statistical analyses of the present study were performed using RStudio, Version 1.1.383.

First, descriptive statistics for both the BISS-A32 and the BTSS were calculated as displayed in Table 3.6.

Table 3.6. Descriptive statistics of BISS-A32 and BTSS scores

	Range	*M*	*(%)*	*SD*	*(%)*
BISS-A32					
Free recall	7.00 - 45.50	27.21	(55.52)	8.23	(16.79)
Recognition memory	9.00 - 16.00	13.73	(85.83)	1.63	(10.20)
Yield	0.00 - 16.00	6.19	(38.69)	4.66	(29.12)
Shift	0.00 - 20.00	6.97	(21.79)	4.00	(12.49)
Free recall	1.50 - 43.50	20.60	(33.22)	9.01	(14.53)
Recognition memory	3.00 - 7.00	6.14	(96.46)	0.82	(14.93)
Yield	0.00 - 16.00	6.18	(37.51)	3.62	(22.35)
Shift	0.00 - 8.00	3.08	(38.47)	2.18	(27.21)

Then, Pearson's product-moment correlations between the BISS-A32 and the BTSS scores were conducted. The results showed a strong correlation between free recall ($r = .60$, $p < .001$) and yield ($r = .69$, $p < .001$) measures. Shift scores ($r = .37$, $p < .001$) and recognition memory scores ($r = .21$, $p = .011$) were significantly associated, but the correlation was not strong.

Finally, the Bland and Altman's (2010) method was applied. To compare the BISS and the BTSS scores, all the measures have been considered as percentages. As shown in Figure 3.7, 3.8, 3.9 and 3.10, most of the subjects fell into the 95% confidence interval depicted by the upper and lower lines for all the measures considered.

Figure 3.7. Plot of differences between BISS-A32 and BTSS in free recall

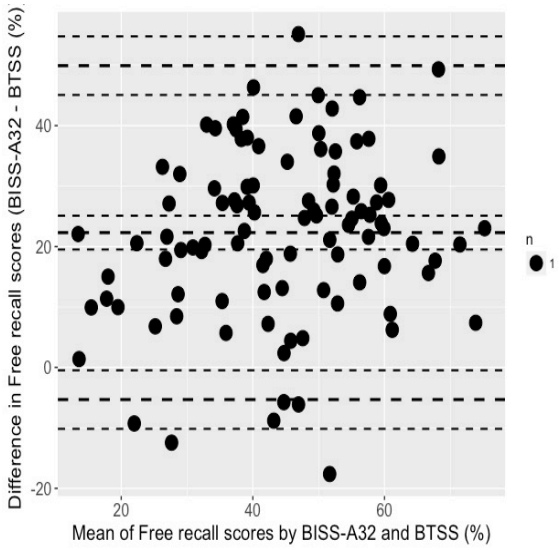


Figure 3.8. Plot of differences between BISS-A32 and BTSS in recognition memory

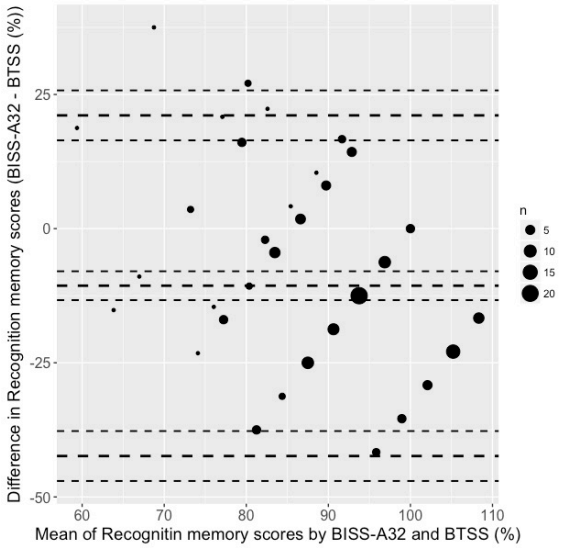


Figure 3.9. Plot of differences between BISS-A32 and BTSS in yield scores

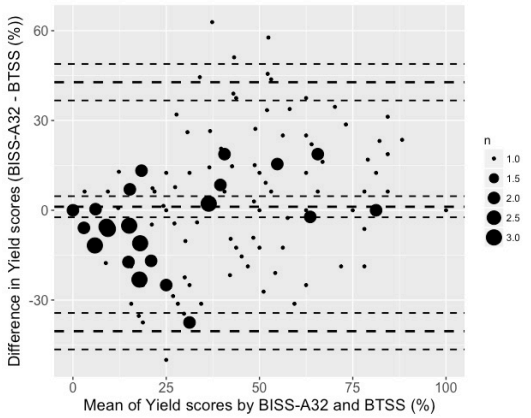
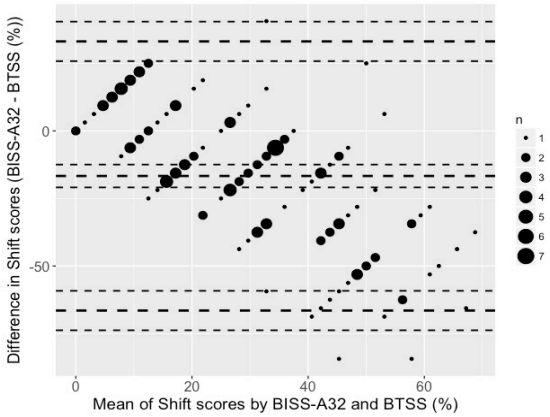


Figure 3.10. Plot of differences between BISS-A32 and BTSS in shift scores



3.3.3. Discussion

The aim of Study 3 was to test the concurrent validity of the BISS-A32 compared with the Italian version of the Bonn Test of Statement Suggestibility (BTSS; Rossi, Benatti, Pesce, Oppo, & Avato, 2011). One-hundred-forty-two primary school children completed the two tests, presented in a counterbalanced order. A significant positive

correlation was found for the free recall score between BISS-A32 and BTSS, indicating good concurrent validity of the measure. It should be noted, however, that even if the BTSS story is longer than the BISS-A32[18] the BISS-A32 story was more easily recalled. Indeed, the children recalled correctly 56% of the BISS story against 33% of the BTSS story. The same result is evident in Figure 3.11, because the difference against mean for each subject tended to be positive.

Regarding the Yield score, a strong positive correlation was found between BISS-A32 and BTSS, indicating good concurrent validity of the measure. The percentage values of the Yield score of both tests resulted to be similar, and the Bland-Altman plot showed enough precision in the measurement of this score when compared to the BTSS. Therefore, when weighted up to another test aimed to assess the same construct, the BISS Yield score revealed satisfactory concurrent validity.

While the yield and free recall scores were highly correlated between the two tests, only a weak correlation was found for shift and recognition memory variables. The weak association between the two shift variables may be explained by considering the different administration of the repeated questions between the two tests. Indeed, the BISS-A32 standard procedure implies that the experimenter reads all the 32 questions and then gives negative feedback to the child to repeat all the questions. Conversely, according to the BTSS standard procedure (Endres, 1997), only 8 (of 31) questions are repeated, and this happens immediately after the child answers to the first question. Therefore, the child would perceive a greater interpersonal pressure when the BTSS questions are repeated rather than the BISS-A32 questions. This would also account for the low rate of BISS-A32 changes made by children. The same result was found in the study of Roma, Sabatello, Verrastro, and Ferracuti (2011) comparing the BTSS with the Gudjonsson Suggestibility Scale.

[18] The BTSS story is composed of 330 words grouped in about 62 informative units while the BISS-A32 is composed of 170 words grouped in 49 informative units.

3.4. STUDY 4. The factor structure of the Bicocca Suggestibility Scales

After having explored the reliability of the short versions of the Bicocca Suggestibility Scales battery, Study 4 aimed to "dig deeper" the pattern of responses to the overall questioning phase and to assess the factor structure of the long form of the instrument.

To achieve this aim, in this study all the possible responses were considered. Indeed, the overall child's performance was considered, by classifying all the combinations of answers to the first and the second questions administration simultaneously. This choice has been made to better understand children's tendency to react to the questions repetition and negative feedback. When questions are repeated children may change their answers. Children's inconsistencies may be used to discredit their testimony. Therefore, a deeper understanding of this phenomenon could help improving jurors' perception of child witness credibility (Berman, Narby, & Cutler, 1995). There could be several reasons underlying this behaviour (Walker, 2013): first, as a general rule of conversation, repeating a question implicitly conveys the message that the first answer was not entirely acceptable. Second, uncertainty and lack of experience or confidence in one's own ability to answer correctly may lead the child to doubt the first answer was right (Gelman, Meck, & Merkin, 1986; Siegal, Waters, & Dinwiddy, 1988). Third, even if children knew that their answer was correct, they could change it, because of incorrect perceptions about a questioner's intentions. Similarly, when the task is perceived as boring, the child may change the answer to the one the adult seems to want to bring the task to a speedy conclusion (Siegal et al., 1988). Finally, the child may not have understood the question the first time it was asked (Walker, 2013).

Moreover, to consider all the children's answers, in Study 3 a second Yield factor was considered, in line with the Gudjonsson Suggestibility Scales model (Bianco & Curci, 2015; Gudjonsson, 1992). This score refers to the children's tendency to provide

the wrong answer to suggestive questions the second time these questions were asked. Therefore, the children's answers to suggestive questions were distinguished according to the administration time in Yield 1 and Yield 2 scores.

3.4.1. Method

Participants

Participants included 331 children (161 males, 170 females) enrolled in 8 primary schools of Northern Italy[19]. They were aged between 6 and 10 year (M_{age} = 8.04, *SD* = 1.16).

Procedure and instruments

The BISS-A, long version, was used. The following measures were considered:

- *Recognition memory* = the number of correct responses to the unbiased yes/no and forced-choice questions during the administration of the first series of questions (maximum = 16).
- *Yield 1* = the number of misleading questions to which the child has given the wrong answer during the administration of the first series of questions (maximum = 16).
- *Yield 2* = the number of misleading questions to which the child has given the wrong answer when the questions were repeated (maximum = 16).
- *Shift* = the number of questions to which the child has changed his or her response after the interviewer's negative feedback (maximum = 32).

By keeping simultaneously in account the children's behaviour during the first and the second administration of the questioning protocol about the story, five different

[19] Sara Messineo, Elisa Piuri, Eleonora Ubertis Albano, Greta Lonati, Elena Casula, Elisa Perosi, Laura Balduchelli, & Debora Smecca collected the data.

coding categories were possible: *Resistance*, *Persistence*, *Negative Shift*, *Positive Shift*, and *Neutral Shift*. A definition of these terms will be provided with a practical example. For instance, the question "Was mum cooking or ironing?" is misleading because the story does not refer to what the mother was doing. Therefore, there are three possible answers: "cooking" and "ironing" are equally wrong, while "I do not know" is the correct option.

- *Resistance*: when the child gave the correct answer to both the presentations of the same question, accurately identifying the right answer to the recognition task and resisting to the suggestion of the misleading questions (e.g. the child says "I don't know" both times).
- *Persistence*: when the child gave the same wrong answer to both the presentations of the same question (e.g. the child says twice "cooking" or "ironing").
- *Negative shift*: when the child firstly chose the correct answer, but changed his or her answer after the interviewer's negative feedback by selecting one wrong option (e.g. the child says "I don't know" the first time and then "cooking" or "ironing").
- *Positive shift*: when the child provided a wrong answer during the first administration of the questioning protocol and corrected himself or herself after the interviewer's negative feedback (e.g. the child says "cooking" or "ironing" the first time and then he or she says "I don't know").
- *Neutral shift*: when the child provided a wrong answer during the first administration of the questioning protocol and changed his or her answer after the interviewer's negative feedback, by selecting the other wrong alternative suggested by the interviewer (e.g. the child once says "cooking" and once says "ironing").

3.4.2. Results

The statistical analyses of the present study were performed using RStudio software, version 1.1.383. The *psych* (Revelle, 2017, 2017) and the *lavaan* packages were used (Rosseel, 2012).

Preliminary analysis was conducted to assess whether there were any differences in the memory and suggestibility scores according to the child's gender. Independent-samples t-tests were performed. The results showed that there were no statistical significative differences between males and females (all the *p*s > .05). Therefore, the subsequent analysis were conducted on the collapsed data.

Then, the number of correct answers per item provided by the children were observed, first by considering separately the two questions administration times (Table 3.7), and after by examining the whole pattern of answers by collapsing the first and the second administration time (Table 3.8). Table 3.9 displays the descriptive statistics of the five possible pattern of answers when considering the two administration times together.

Table 3.7. Number of correct responses per item for the first and the repeated administration and number of Shifts (N=331 children) – in brackets the percentage of children

		First administration		Repeated questioning		Shift	
ITEM1	(control)	321	(97.0)	281	(84.9)	44	(13.3)
ITEM2	(control)	260	(78.5)	211	(63.7)	85	(25.7)
ITEM3	(control)	312	(94.3)	250	(75.5)	76	(23.0)
ITEM4	(suggestive)	174	(52.6)	145	(43.8)	85	(25.7)
ITEM5	(suggestive)	273	(82.5)	227	(68.6)	81	(24.5)
ITEM6	(control)	232	(70.1)	185	(55.9)	85	(25.7)
ITEM7	(suggestive)	255	(77.0)	194	(58.6)	83	(25.1)
ITEM8	(control)	320	(96.7)	317	(95.8)	7	(2.10)
ITEM9	(suggestive)	258	(77.9)	220	(66.5)	66	(19.9)
ITEM10	(suggestive)	218	(65.9)	169	(51.1)	67	(20.2)
ITEM11	(control)	324	(97.9)	319	(96.4)	10	(3.00)
ITEM12	(control)	302	(91.2)	192	(58.0)	123	(37.2)
ITEM13	(control)	318	(96.1)	315	(95.2)	9	(2.70)
ITEM14	(control)	223	(67.4)	203	(61.3)	55	(16.6)
ITEM15	(suggestive)	196	(59.2)	161	(48.6)	107	(32.3)
ITEM16	(control)	310	(93.7)	305	(92.1)	12	(3.60)
ITEM17	(control)	216	(65.3)	115	(34.7)	169	(51.1)
ITEM18	(suggestive)	239	(72.2)	226	(68.3)	50	(15.1)
ITEM19	(suggestive)	149	(45.0)	137	(41.4)	42	(12.7)
ITEM20	(control)	280	(84.6)	225	(68.0)	87	(26.3)
ITEM21	(suggestive)	184	(55.6)	162	(48.9)	72	(21.8)
ITEM22	(suggestive)	284	(85.8)	267	(80.7)	29	(8.80)
ITEM23	(control)	271	(81.9)	254	(76.7)	44	(13.3)
ITEM24	(control)	249	(75.2)	186	(56.2)	107	(32.3)
ITEM25	(suggestive)	269	(81.3)	251	(75.8)	52	(15.7)
ITEM26	(suggestive)	211	(63.7)	177	(53.5)	64	(19.3)
ITEM27	(suggestive)	251	(75.8)	244	(73.7)	81	(24.5)
ITEM28	(suggestive)	169	(51.1)	156	(47.1)	70	(21.1)
ITEM29	(suggestive)	223	(67.4)	218	(65.9)	139	(42.0)
ITEM30	(suggestive)	143	(43.2)	180	(54.4)	88	(26.6)
ITEM31	(control)	219	(66.2)	165	(49.8)	111	(33.5)
ITEM32	(control)	276	(83.4)	267	(80.7)	28	(8.50)

Table 3.8. Response per item (N=331 children).

		Resistance		Persistence		Negative shift		Positive shift		Neutral shift	
ITEM1	(control)	280	(84.6)	7	(2.11)	40	(12.1)	2	(0.60)	2	(0.60)
ITEM2	(control)	211	(63.7)	35	(10.6)	49	(14.8)	31	(9.37)	5	(1.51)
ITEM3	(control)	250	(75.5)	5	(1.50)	62	(18.7)	11	(3.32)	3	(0.91)
ITEM4	(suggestive)	117	(35.3)	129	(39.0)	57	(17.2)	28	(8.46)	0	(0.00)
ITEM5	(suggestive)	221	(66.8)	29	(8.76)	52	(15.7)	6	(1.81)	23	(6.95)
ITEM6	(control)	185	(55.9)	61	(18.4)	47	(14.2)	27	(8.16)	11	(3.32)
ITEM7	(suggestive)	183	(55.3)	65	(19.6)	72	(21.8)	11	(3.32)	0	(0.00)
ITEM8	(control)	317	(95.8)	7	(2.11)	3	(0.91)	4	(1.21)	0	(0.00)
ITEM9	(suggestive)	213	(64.4)	52	(15.7)	45	(13.6)	7	(2.11)	14	(4.23)
ITEM10	(suggestive)	160	(48.3)	104	(31.4)	58	(17.5)	9	(2.72)	0	(0.00)
ITEM11	(control)	319	(96.4)	2	(0.60)	5	(1.51)	4	(1.21)	1	(0.30)
ITEM12	(control)	192	(58.0)	16	(4.83)	110	(33.2)	8	(2.42)	5	(1.51)
ITEM13	(control)	315	(95.2)	7	(2.11)	3	(0.91)	5	(1.51)	1	(0.30)
ITEM14	(control)	203	(61.3)	73	(22.1)	20	(6.04)	28	(8.46)	7	(2.11)
ITEM15	(suggestive)	142	(42.9)	82	(24.8)	54	(16.3)	19	(5.74)	34	(10.3)
ITEM16	(control)	305	(92.1)	14	(4.23)	5	(1.51)	7	(2.11)	0	(0.00)
ITEM17	(control)	115	(34.7)	47	(14.2)	101	(30.5)	54	(16.3)	14	(4.23)
ITEM18	(suggestive)	213	(64.4)	68	(20.5)	26	(7.85)	13	(3.93)	11	(3.32)
ITEM19	(suggestive)	122	(36.9)	167	(50.5)	27	(8.15)	15	(4.53)	0	(0.00)
ITEM20	(control)	225	(68.0)	19	(5.74)	55	(16.6)	19	(5.74)	13	(3.93)
ITEM21	(suggestive)	149	(45.0)	110	(33.2)	35	(10.6)	13	(3.93)	24	(7.25)
ITEM22	(suggestive)	261	(78.9)	41	(12.4)	23	(6.95)	6	(1.81)	0	(0.00)
ITEM23	(control)	254	(76.7)	33	(10.0)	17	(5.14)	21	(6.34)	6	(1.81)
ITEM24	(control)	186	(56.2)	38	(11.5)	63	(19.0)	36	(10.9)	8	(2.42)
ITEM25	(suggestive)	234	(70.7)	45	(13.6)	35	(10.6)	17	(5.14)	0	(0.00)
ITEM26	(suggestive)	171	(51.7)	96	(29.0)	40	(12.1)	6	(1.81)	18	(5.44)
ITEM27	(suggestive)	207	(62.5)	43	(13.0)	44	(13.3)	37	(11.2)	0	(0.00)
ITEM28	(suggestive)	146	(44.1)	115	(34.7)	23	(6.95)	10	(3.02)	37	(11.2)
ITEM29	(suggestive)	151	(45.6)	41	(12.4)	72	(21.8)	67	(20.2)	0	(0.00)
ITEM30	(suggestive)	132	(39.9)	111	(33.5)	11	(3.32)	48	(14.5)	29	(8.76)
ITEM31	(control)	165	(49.8)	55	(16.6)	54	(16.3)	43	(13.0)	14	(4.23)
ITEM32	(control)	267	(80.7)	36	(10.9)	9	(2.72)	14	(4.23)	5	(1.51)

In brackets the percentage of children who chose that answer on the total sample.

Table 3.9. Descriptive statistics of the responses (N = 331 children, 32 items)

	Range	M	SD	Skewness	Kurtosis
Resistance	3 - 32	19.97	6.48	-0.42	-0.81
Persistence	0 - 19	5.30	3.91	0.84	0.13
Negative shift	0 - 16	3.98	2.87	0.93	0.89
Positive shift	0 - 10	1.89	1.60	1.38	3.12
Neutral shift	0 - 7	0.86	1.30	2.04	4.57

Figure 3.11 shows the distribution of the different types of Shift basing on the direction of the change made by the child. Similarly to Study 1 and Study 2, the Negative Shift was the predominant direction when changing answer after the negative feedback.

Figure 3.11. Percentage of the different types of shift on the total changes made

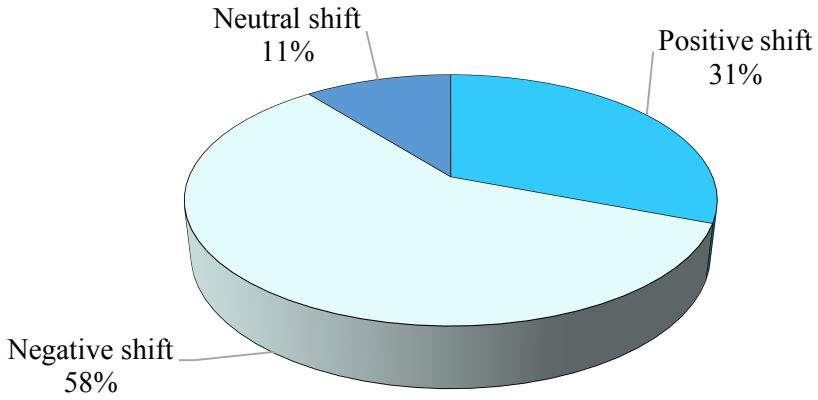


Table 3.10. Descriptive information for memory and suggestibility measures

	Range	M	SD	Skewness	Kurtosis
Free recall	4.00 - 45.5	26.11	8.52	-0.52	-0.21
Recognition memory	5.00 - 16.0	13.39	1.91	-1.10	1.57
Yield 1	0.00 - 16.0	5.44	4.56	0.60	-0.79
Yield 2	0.00 - 16.0	6.53	5.04	0.29	-1.28
Shift	0.00 - 20.0	6.73	4.02	0.77	0.51
Total suggestibility	0.00 - 35.0	12.17	7.51	0.42	-0.68

Table 3.10 displays the descriptive statistics of the memory and suggestibility scores of the BISS-A32. Similarly to the results of Study 1 and Study 2 with the short versions of the BISS, the recognition memory score was asymptotic.

Then, internal consistency was assessed by using ordinal polychoric correlation-based alphas, which provide a better estimate of reliability than Cronbach's alpha for binary and ordinal response scales (Gadermann, Guhn, & Zumbo, 2012). The 32 items were modelled as categorical variables, by considering the answer provided by the child as "correct" or "wrong". Ordinal alphas were computed to assess the internal consistency of recognition memory, yield 1, yield 2 and shift scales. Table 3.11, 3.12, 3.13 and 3.14 display polychoric item-total correlation coefficients corrected for item overlap and scale reliability, as well as the reliability if the item is dropped. Polychoric correlation matrix was considered because the polychoric correlation estimates the relationship between the dichotomous variables more accurately than Pearson's correlation does (Gadermann et al., 2012). The results indicate that yield 1 (16 items; *ordinal* $\alpha = 0.95$), yield 2 (16 items; *ordinal* $\alpha = 0.96$) and total suggestibility (48 items; *ordinal* $\alpha = 0.94$) scales have a high internal consistency. The internal consistency for the shift scale (32 items; *ordinal* $\alpha = 0.85$) is acceptable, while the recognition memory scale (16 items; *ordinal* $\alpha = 0.7$) cannot be considered reliable enough. The low values of the recognition memory polychoric item-total correlation coefficients may be ascribed to the lack of variability due to the ceiling effect. A Principal Component Analysis was conducted to explore the factor structure underlying the BISS scores[20] and the results showed that it was not possible to distinguish the recognition memory factor.

[20] For the sake of brevity, the results of this analysis will not be reported here.

Table 3.11. Item-total correlations and reliability for the recognition memory factor.

Recognition memory	Polychoric correlation	Reliability if the item is dropped
Item 1	.26	.71
Item 2	.47	.68
Item 3	.54	.67
Item 6	.19	.72
Item 8	.36	.70
Item 11	.64	.66
Item 12	.30	.70
Item 13	.49	.68
Item 14	.54	.67
Item 16	.23	.71
Item 17	.57	.67
Item 20	.42	.69
Item 23	.37	.69
Item 24	.37	.70
Item 31	.55	.67
Item 32	.55	.67

Table 3.12. Item-total correlations and reliability for the yield 1 factor.

Yield 1	Polychoric correlation	Reliability if the item is dropped
Item 4	.59	.96
Item 5	.75	.95
Item7	.42	.96
Item 9	.83	.95
Item 10	.82	.95
Item 15	.80	.95
Item 18	.77	.95
Item 19	.85	.95
Item 21	.83	.95
Item 22	.86	.95
Item 25	.61	.95
Item 26	.89	.95
Item 27	.87	.95
Item 28	.84	.95
Item 29	.77	.95
Item 30	.82	.95

Table 3.13. Item-total correlations and reliability for the shift factor.

Shift	Polychoric correlation	Reliability if the item is dropped
Item 1	.33	.85
Item 2	.38	.84
Item 3	.50	.84
Item 4	.27	.85
Item 5	.68	.84
Item 6	.32	.85
Item 7	.43	.84
Item 8	.40	.84
Item 9	.55	.84
Item 10	.26	.85
Item 11	.67	.84
Item 12	.44	.84
Item 13	.36	.85
Item 14	.47	.84
Item 15	.49	.84
Item 16	.27	.85
Item 17	.25	.85
Item 18	.47	.84
Item 19	.12	.85
Item 20	.55	.84
Item 21	.48	.84
Item 22	.59	.84
Item 23	.35	.85
Item 24	.43	.84
Item 25	.53	.84
Item 26	.63	.84
Item 27	.41	.84
Item 28	.41	.84
Item 29	.26	.85
Item 30	.36	.85
Item 31	.15	.85
Item 32	.57	.84

Table 3.14. Item-total correlations and reliability for the yield 2 factor.

Yield 2	Polychoric correlation	Reliability if the item is dropped
Item 4	.78	.96
Item 5	.84	.96
Item7	.66	.96
Item 9	.79	.96
Item 10	.82	.96
Item 15	.83	.96
Item 18	.78	.96
Item 19	.80	.96
Item 21	.83	.96
Item 22	.86	.96
Item 25	.69	.96
Item 26	.93	.96
Item 27	.78	.96
Item 28	.91	.96
Item 29	.57	.96
Item 30	.85	.96

Then, a Confirmatory Factor Analysis (CFA) was run. A first model including the recognition memory items was tried, but it did not converge. Therefore, after considering the ceiling effect, the reliability analysis results and the Principal Component Analysis results, it seemed appropriate to not include the recognition memory scale further in the model. Moreover, the shift scale was not suitable to be included in the factorial model because of its distribution and the high presence of constants. Indeed, when examining every item singularly, it appeared that children rarely changed their first answer, nevertheless of the direction of the change.

Thus, two different CFA models were examined. First, one-factor structure model was tested by entering the Yield 1 items only. Subsequently a two-factors model was tested, by including the Yield 1 and the Yield 2 items. While Yield 1 refers to the extent to which the children were affected by suggestive questions the first time these were asked, Yield 2 reflects the children's reaction to suggestive questions when these are

repeated. As shown in Table 3.15, the following goodness-of-fit indices were considered for the one- and two- factor CFA models of BISS-A32: Chi-square, Root Mean Square Residual (RMR), Standardized Root Mean Square Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), Non-Normed Fit Index (NNFI) and the Bentler's Comparative Fit Index (CFI). According to Hu & Bentler's (1999) cut-off criteria, the 1-factor model showed a good fit, while the two-factor model was not acceptable.

Table 3.15. Goodness-of-fit indices for the one- and two- factor CFA models of BISS-A32.

	Chi-square	do	p	RMR	SRMR	RMSEA	NNFI	CFI
1-factor solution	221.29	104	.00	.01	.05	.06	.93	.94
2-factor solution	2096.71	463	.00	.01	.07	.10	.70	.72

Table 3.16. Standardised loadings of the one-factor model of the BISS-A32

Item	Yield 1
4	.43
5	.52
7	.25
9	.62
10	.65
15	.64
18	.61
19	.67
21	.68
22	.59
25	.41
26	.77
27	.68
28	.70
29	.62
30	.64

Moreover, as shown in Table 3.16, standardised factor loadings for the one-factor model were above .30, except item 7.

Pearson's correlations were then conducted to assess the association with age for all the variables considered. Age was associated with an improvement in the free recall performance (r = .425, p < .001) and a reduction of the tendency to Yield to suggestive questions (r = -.522, p < .001), to change answers after receiving negative feedback (r = -.405, p < .001), and therefore, the total suggestibility score (r = -.534, p < .001). When considering the global children's performance, a positive association emerged between age and the ability to resist to suggestion (r = .547,

$p < .001$); conversely, the younger children were more likely to persist in the errors ($r = -.489$, $p < .001$) than the older ones.

The different trends of Shifting answer were then included in the correlational analysis. The absolute scores were negatively related with age (Positive Shift: $r = -.125$, $p = .023$; Negative Shift: $r = -.376$, $p < .001$; Neutral Shift: $r = -.271$, $p < .001$), by reflecting a reduction with age in the propensity to be inconsistent when questions were repeated. However, when observing the proportion of each type of Shift on the total Shift score, a different picture emerged. Indeed, the relationship between Negative Shift and Neutral Shift with age was still negative (Negative Shift: $r = -.147$, $p = .008$; Neutral Shift: $r = -.133$, $p = .016$), whereas a positive association between age and Positive Shift was found ($r = .235$, $p < .001$).

Finally, the post-event misinformation effect was considered. Figure 3.12 displays the percentage of children who presented or did not the misinformation effect. Figure 3.13 shows the number of times the children included an element suggested during the questioning phase in their subsequent recall of the story.

A t-test was conducted to examine whether there were age differences between the children who manifested the misinformation effect and those who did not. The results showed a significant difference of age ($t_{(329)} = 4.12$, $p < .001$, $d = .49$): the children who displays the misinformation effect ($M_{age} = 8.23$, $SD = 1.13$) were younger than the other ones ($M_{age} = 7.70$, $SD = 1.15$).

Figure 3.12. Presence or absence of the misinformation effect

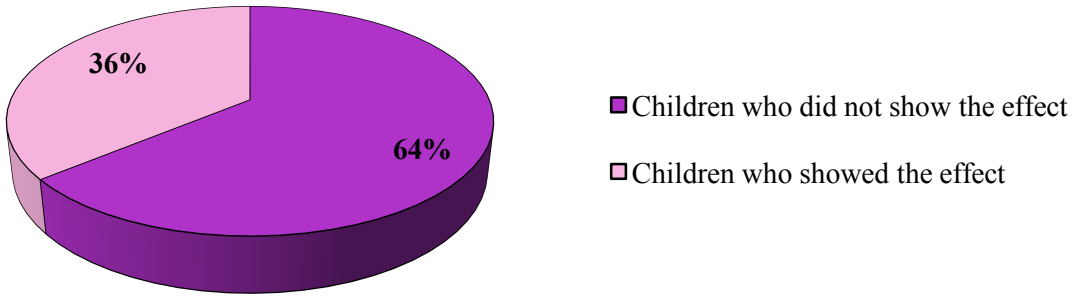
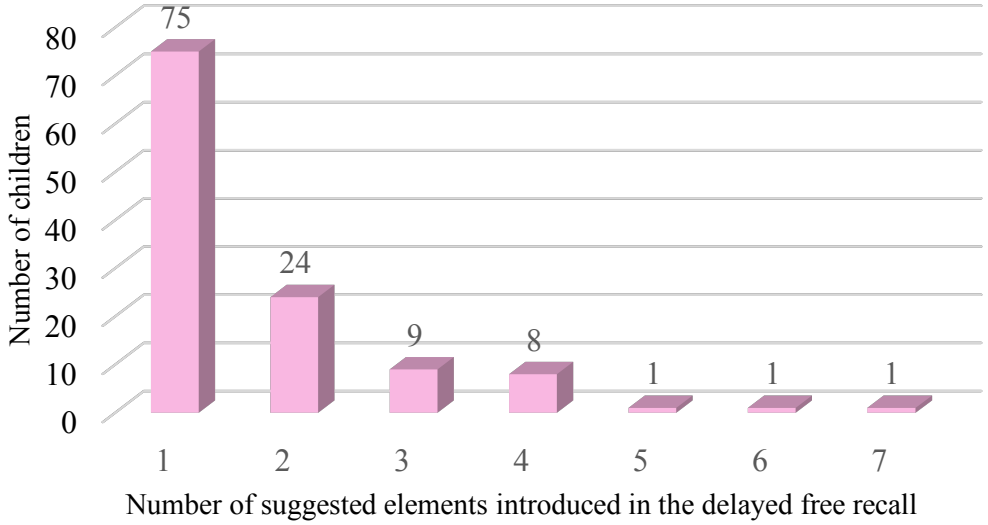


Figure 3.13. Frequency of the misinformation effect phenomenon



3.4.3. Discussion

Study 4 aimed to assess the factor structure of the long form of the BISS. First, internal consistency analysis was conducted and the results replicated what already found in Study 1 and Study 2 with the short versions of the instrument. Therefore, the reliability of the Yield and the Shift scores has been consistently proved, while the recognition memory score had poor internal consistency for this factor. It may be assumed that the control questions, i.e. those specific-closed questions about elements actually present in the story, were too easy for school-aged children, so that a ceiling effect was repeatedly observed. This lack of variability interfered with all the analysis, by mining the reliability of the measure. Future research should be conducted to solve this limit of the instrument. A possible reason for this unexpected ceiling effect could be that specific-closed questions are per se someway suggestive (Endres, 1997). Therefore, when asking children to answer to this kind of question for elements actually presented in the story after a very short time, the task would be very easy. Moreover, by limiting their possibility of answer

to two options, there would be an increased probability of chance responding. This possibility should be considered in future studies.

The CFA tested the one-factor and two-factor model solutions. The results supported the model with the Yield factor, while the results of the model comprising the Yield 2 factor were not satisfactory enough to justify the introduction of this second scale in the BISS procedure.

Additionally, in Study 4 children's behaviour was examined by keeping simultaneously in account the answer to the first and the second administration of the questioning protocol about the story. It resulted that the most dominant pattern of responses was "resistance", i.e. the tendency to give the correct answer and then maintain this answer even after the adult's negative feedback. Conversely, the "persistence" tendency revealed the obstinacy of the child in providing the same wrong answer both the first time the question was proposed and the second one. Although rare, this tendency affected the control question in addition to misleading questions. This finding should be a warning that sometimes children consistently report wrong information even when they should know the correct answer.

Regarding the Shift score, to date, it has been considered as a simple composite factor. This score has been proposed by Gudjonsson, who initially proposed to compute this measure by summing the suggestive questions only (Gudjonsson, 1983, 1984). Later, in order to overtake some limitations due to the poor reliability of this factor, (Singh & Gudjonsson, 1987) recommended using all the 20 questions of the Gudjonsson Suggestibility Scale (GSS) to obtain the Shift score. Gignac and Powell (2009) criticised the composite Shift score and proposed the distinction in "Positive" and "Negative" Shift, according to the direction of the change made by the person. However, their analysis did not confirm their hypothesis and they then suggested to limit the use of the GSS to the Yield scale only. Similarly to the Video Suggestibilty Scale for Children (VSSC), the GSS Shift score could have been affected by an "acquiescence bias" that could render

more difficult to shift from “yes” to “no” (McFarlane & Powell, 2002). The BISS-A32 comprises the same number of control questions requiring a “yes” rather than a “no” answer and this would keep under control that kind of response bias. The distinction made with the BISS procedure in Positive, Negative and Neutral Shift would add information about this complex phenomenon. It is interesting that, while the general tendency to react to interpersonal pressure by changing the initial answer was negatively associated with age, when separating the directions of change different trends emerged. Indeed, the proportions on the total shift score revealed a reduction of Negative and Neutral Shift with increasing age and a concurrent increase in the Positive Shift. This latter measure would reflect a positive reaction for the child, the same used in everyday interaction to enhance learning (Binet, 1900). Conversely, Negative Shift and Neutral Shift would mean a worsening in accuracy rate as a consequence of interpersonal pressure. This issue could have great applied value and deserves further examination.

Finally, post-event misinformation was examined. The results showed that 36% of the children introduced at least a suggested element in the delayed recall of the story, and one child reported seven suggested details too. Younger children displayed the phenomenon more often than the older ones. This finding provide support to the use of the misinformation effect measure in the BISS procedure, because it showed that some children are likely to incorporate suggested information in their subsequent recall even after a short delay. This procedure differs from the standard misinformation paradigm tested with a recognition task (Loftus, 2005). Future studies should address the predictive validity of this measure for personal-experienced events.

CHAPTER 4

Individual differences in children's suggestibility: the role of cognitive functioning.[21]

Despite the documented age-related differences, children within a given age group vary in terms of their memory performances, in the occurrence of false memories and their susceptibility to suggestion (Chae & Ceci, 2005). Given the strong interest in the causes of this variability, many studies have explored the individual characteristics of children that may moderate their vulnerability to suggestion.

A significant amount of research has documented a wide range of cognitive and temperamental factors that have been found to mediate differences in suggestibility in children. However, the patterns of findings in relation to these variables have not been consistent (Bruck & Melnyk, 2004). Within the area of cognitive factors, the most relevant are memory, intelligence and executive functioning.

It has been shown that poor general memory leads to an increased susceptibility to false memories (Endres, 1997; McFarlane et al., 2002; Warren et al., 1991), but this relationship has not been consistently observed (Chae & Ceci, 2005; Danielsdottir et al., 1993; Roebers & Schneider, 2002). Furthermore, it is not clear which aspects of memory processes are involved in suggestibility and there is considerable debate concerning this issue (Ceci & Bruck, 1993). Few studies have included an independent assessment of the memory ability of children, and generally they have not detected relationships with

[21] A slightly modified version of this study has been published as:

Caprin, C., Benedan, L., Ciaccia, D., Mazza, E., Messineo, S., & Piuri, E. (2016). True and false memories in middle childhood: the relationship with cognitive functioning. *Psychology Crime & Law, 22(5)*, 473–494. https://doi.org/10.1080/1068316X.2016.1168420.

suggestibility, so it is not clear whether memory ability should be considered a useful predictor of suggestibility (Bruck & Melnyk, 2004).

Regarding intelligence, the question of whether individuals with lower IQs are more suggestible than individuals with higher IQs is of great relevance to forensic cases, but despite the large number of studies, the relationship between intelligence (or related constructs) and suggestibility remains controversial, both in adults and children (Bruck, Ceci, & Melnyk, 1997; Zhu et al., 2010). Although intelligence is a complex ability, discrepancies do not depend on the particular characteristic investigated, because some studies have found significant correlations with both verbal and non-verbal IQs (McFarlane et al., 2002; Roma et al., 2011; Zhu et al., 2010), whereas others have not found any significant relationship (Bruck et al., 1997; Bruck & Melnyk, 2004). Various authors have suggested that only below-average intelligence correlates significantly with suggestibility (Bruck et al., 1997; Henry & Gudjonsson, 2003, 2007).

Another construct that has been associated to suggestibility is executive functioning. This term refers to a broad category of skills that underlie goal-directed behaviour, such as: planning, executive control, inhibitory control of response, sustained and selective attention, attentive flexibility and working-memory (Zelazo, Craik, & Booth, 2004). It should be underlined that although some authors consider executive functioning and working memory separately (Roebers & Schneider, 2005a), these concepts overlap. In the present study, attention and working memory function will be considered together, because most neuropsychological tests give mixed measures of working memory and attentive abilities. Research on the correlates of interrogative suggestibility have found that high levels of inhibition reduce the tendency of children to confirm false events and to accept misleading information (Alexander et al., 2002; Melinder, Endestad, & Magnussen, 2006; Roberts & Powell, 2005; Schaaf, Alexander, & Goodman, 2008). In addition, studies with adults have found a relationship between propensity to integrate misleading post-event information into recall and working

memory capacity (Zhu et al., 2010), therefore age differences in suggestibility might be due to a shorter memory span in younger children than in older ones (Finnilä et al., 2003).

Most of the previous studies that have examined developmental changes in memory performance and suggestibility have addressed preschoolers, because this age group is the most vulnerable to suggestion (Ceci & Bruck, 1993). However, the suggestibility levels remain high throughout elementary school years and it would be of interest to examine age-related and individual differences during this period of life. Indeed, the psychological factors which underlie suggestibility in preschool children might differ from those underlying suggestibility in older children (Bruck et al., 1997).

4.1. STUDY 5. The relationship between cognitive functioning and interrogative suggestibility

Study 5 aimed to fill the gap by investigating the relationship between memory and suggestibility with cognitive functioning in school-aged children. The findings of previous studies are inconsistent and contradictory (Bruck & Melnyk, 2004). The hypothesis is that executive functioning may be related to some aspects of confabulations and suggestibility, since both seem to depend on the frontal lobe maturation, while intelligence, measured in a way that excludes the influence of verbal ability, should be unrelated.

Specifically the free-narrative account of a story will be measured, as measure of the amount of true information recalled by children; moreover, confabulations during free recall of the story, such as Distortions and Fabrications (Gudjonsson & Clare, 1995) will be considered. Investigation of the presence of confabulations can be significant from the point of view of forensic psychology, since the presence of completely fictional false elements, albeit unaware, must be taken into account when interviewing a child, as well as the potential consequences.

Suggestibility is a multidimensional phenomenon but, generally, studies have addressed interrogative suggestibility or the misinformation effect separately, obscuring possible relationships between them. It was therefore of interest to investigate interrogative suggestibility and the subsequent misinformation effect in the same study to better understand the relationship between the different forms of suggestibility.

Finally, this study aimed to extend knowledge relative to the complex relationship between memory and suggestibility.

4.1.1. Method

Participants

This study took place at 9 primary schools in Milan (Italy)[22]. The participants were 372 children, (174 males and 198 females) ranging in age from 6.01 to 11.29 years (M_{age} = 8.61, *SD* = 1.21). Informed written parental consent was obtained for all the children participating in the study. The children were divided into two age groups: G1=193 from 6 to 8.5 (M_{age} = 7.63, *SD* = .52), G2=179 from 8.5 to 11 (M_{age} = 9.67, SD = .76). Approximately the same number of males and females comprised each age group. The children were predominantly of Italian ethnic heritage (89.1%) and came from low (24.4%), middle (53.2%) and high (20.5%) SES families. The SES was obtained by considering parent education level and occupation.

Instruments

The BISS-A, long version, was used to assess memory and suggestibility. The event-memory quantity was measured by *Free Recall* score. It was derived from the children's immediate free-narrative account of the story and represented the amount of information that the children recounted spontaneously after the story had been read out.

[22] Donatella Ciaccia, Eleonora Mazza, Silvia Radaelli, Sara Messineo, Elisa Piuri, Eleonora Ubertis Albano, Sonia Mattavelli, Greta Lonati, Elena Casula, and Flavia Achermann.

The story was segmented into 44 information units and children received one point for each unit correctly recalled and half a point for each one partially recalled. The range of possible scores was 0-49.

The *confabulations* were coded from the immediate free recall transcript. In line with Gudjonsson, the confabulations were categorized as *Distortions* or *Fabrications* (Clare et al., 1994). A distortion was scored each time the recall of the story contained either a major change in the details of an existing idea or substitution an idea from one part to the story with another (for example, to attribute to a character an action performed by another one). A fabrication was scored each time the subjects introduced an entirely novel element that was neither mentioned nor implied by the story (for example to attribute to a character an action that had not occurred in the story). Finally, the *Confabulation* score was computed by summing the distortions and fabrications.

The following measures of suggestibility were obtained from the interview: *Yield* and *Shift*. The Yield score was obtained by adding the number of wrong answers to misleading questions (e.g. by answering to a "yes/no" question or choosing one of a pair of false alternatives when the right answer was "I don't know") prior to having been given negative feedback. The range of possible scores was 0-16. The Shift score indicated the extent to which the children made changes their answers under pressure of negative feedback. The range of possible scores was 0-32 and these were obtained during the repeated-interview only. Finally, a Total Suggestibility was computed by summing the Yield and Shift scores and, therefore, the range of possible scores was 0-48.

The measure concerning the *post-event misinformation effect* was obtained by assigning one point each time one of the pieces of information suggested during the interview was included in the delayed free-narrative account.

Cognitive functioning was assessed in terms of intelligence and executive functions. Intelligence was assessed by means of Raven's Coloured Progressive Matrices

(CPM; Raven, 1998), a commonly employed test to measure fluid intelligence, suitable for children aged between 6 and 11 years.

Three executive functions components were considered: *working memory* was assessed by means of the Digit Span Test of the WISC-III (Wechsler, 1991). It consists of two subtests: forward and backward digit recall. Sequences of digits of increasing length are read aloud. After every sequence the children are required to repeat the digits either forwards or backwards. Children have to recall two trials correctly in order to continue with the test. The number of correctly reproduced strings is considered to be a measure of simple verbal working memory.

Attention was assessed by means of the Test of Auditory Attention (Test di Attenzione Uditiva –TAU). This test requires the subject to mentally count a series of sounds, with each trial consisting of a different number of sounds. The test measures working memory functioning (capacity and phonological loop) and sustained attention.

Inhibitory Control was assessed by means of the Go-No Go Test (Test delle ranette) of the Italian Battery for ADHD (Batteria Italiana per ADHD, Marzocchi, Re, & Cornoldi, 2010). The Go-No Go Test is based on the "Walk Don't Walk Test" (Manly, Robertson, Anderson, & Niparko, 1999). Children are asked to make a dot with a pen in a sequence of 12 cells every time a specific sound is presented and to refrain from doing so, when another sound is presented. The test measures inhibition, sustained and selective attention.

Procedure

The study was presented to the participants as a study about memory. All the children were tested individually in a quiet room at their school.

The raw scores obtained by the children in the cognitive tests were converted into the normative scaled scores, according to the scoring procedures specified in the manuals.

4.1.2. Results

Preliminary analyses showed that memory and suggestibility measures did not differ in terms of gender, so the subsequent analyses were conducted on the collapsed data. Considering the wide age range of the sample, analyses were conducted on the entire sample data and distinguishing two age-groups: G1 (6.01 – 8.50) and G2 (8.51 – 11.29) in order to examine the phenomenon in depth.

The results of the statistical analyses are presented in three major sections: a) age effects on the quantitative and qualitative measures of memory and their relationships with cognitive functioning b) age effect on the suggestibility measures and their relationships with cognitive functioning; c) the relationships among memory and suggestibility measures.

Memory ability: Age effects and relationships with cognitive functioning

The descriptive statistics of the memory measures are shown in Table 4.1.

Table 4.1. Descriptive statistics of free recall and confabulations by age group.

	G1			G2			ENTIRE GROUP		
	N (%)	Range	M (SD)	N (%)	Range	M (SD)	N (%)	Range	M (SD)
Free Recall	193	5-40	22.63	179	10-40	27.85	372	5-42	25.14
	(100)		(7.18)	(100)		(6.18)	(100)		(7.20)
Distortions	143	0-7	1.46	129	0-7	1.48	272	0-7	1.47
	(74)		(1.26)	(73)		(1.37)	(73)		(1.31)
Fabrications	63	0-6	0.51	34	0-4	0.31	97	0-6	.41
	(33)		(0.92)	(19)		(0.76)	(26)		(.85)

Free recall scores showed a normal distribution (skewness: = -.31, kurtosis: free recall = -.30). To check the age effect on quantitative measures, a series of t-tests were conducted on two age groups (G1 vs G2). Significant effects of age were found in free

recall ($t_{(368.053)} = -7.53$, $p < .001$, $d = .78$). As shown in Table 4.1, G2 children's scores were better than those in G1 both in free recall. The confabulations (distortions = D & fabrications = F) did not show a normal distribution (skewness: D = 1.02, F = 2.66; kurtosis: D = 1.45, F = 8.63), in fact only a few children demonstrated the phenomenon, especially fabrications. To check age differences on confabulations (D and F), a series of Mann Whitney U tests were conducted.

The analyses showed no significant differences among the age-groups in the distortions (p=.88), but the groups differed in fabrications ($p = .003$). G1 children made more errors than G2. Furthermore, a Mann-Whitney test revealed that the number of distortions were significantly higher than the number of fabrications ($p < .001$).

Because of the overall results, the subsequent analyses of the entire sample and age-groups were conducted controlling for age. To investigate the relationships between the memory scores and fluid intelligence as well as executive-functions, partial correlations were used. The analyses were conducted on the Entire sample and distinguishing G1 and G2. The descriptive statistics and the correlation values are shown in Table 4.2 and 4.3. As shown in Table 4.2, the Go-No Go and TAU scores were quite high, hence it could be that ceiling effects on these measures reduced the chances of finding a significant relationship.

Table 4.2. Descriptive statistics of the scores concerning the cognitive functioning tests.

	G1		G2		ENTIRE GROUP	
	M	SD	M	SD	M	SD
Intelligence	70.24	24.04	57.09	24.72	63.86	25.21
Working Memory	8.37	3.44	10.16	4.21	9.23	3.93
Attention	62.12	23.23	57.81	24.1	60.05	23.72
Inhibitory Control	63.37	24.86	58.04	24.23	60.79	24.67

In terms of the analyses of the entire sample, all memory measures were significantly correlated. Free recall correlated negatively with distortions and fabrications. In addition, distortions and fabrications were also positively correlated with each other. The analyses of the two age-groups data confirmed that memory measures were significantly correlated: free recall correlated negatively with distortions in G2 and with fabrications in G1. Furthermore, distortions and fabrications were also positively correlated, but in G2 only.

Table 4.3. Partial correlations among memory and cognitive functioning measures.

		1	2	3
G1	1. Free recall			
df = 184	2. Distortions	-.115		
	3. Fabrications	-.154*	.105	
	4. Intelligence	.193**	-.119	-.138
	5. Working Memory	.185*	.169*	-.202**
	6. Attention	.068	-.026	.104
	7. Inhibitory Control	.105	-.1	.167*
G2	1. Free recall			
df = 175	2. Distortions	-.229**		
	3. Fabrications	-.126	.245***	
	4. Intelligence	.113	.016	.002
	5. Working Memory	.046	-.024	.009
	6. Attention	.023	-.014	.061
	7. Inhibitory Control	.075	-.004	.066
entire group	1. Free recall			
df = 362	2. Distortions	-.145**		
	3. Fabrications	-.149**	.147**	
	4. Intelligence	.167***	-.034	-.093
	5. Working Memory	.134*	.092	-.117*
	6. Attention	.048	-.016	.088
	7. Inhibitory Control	.071	-.08	.134**

Regarding the relationships among memory and cognitive measures, the analyses of the entire sample data showed that free recall was positively correlated with working

memory and with intelligence. Fabrications were negatively correlated with working memory and positively with the inhibitory control (Go-No Go scores). The analyses of the age-groups data confirmed that free recall was positively correlated with working memory and free recall with intelligence in G1, while this association did not emerge in G2. In G1, distortions were positively correlated with working memory, while fabrications were negatively correlated with it and positively with the Go-No Go scores. We did not find significant associations between qualitative measures of memory and cognitive functioning in G2.

Suggestibility: age effects and relationships with cognitive functioning

The Descriptive Statistics of the suggestibility measures are shown in Table 4.4.

Table 4.4. Descriptive Statistics of suggestibility measures by age group.

		N	(%)	Range	M	SD
G1	Yield	170	(88.08)	0 - 15	4.68	(3.89)
	Shift	190	(99.45)	0 - 18	7.00	(3.87)
	Total suggestibility	192	(99.48)	0 - 32	12.62	(6.98)
	Misinformation	71	(36.78)	0 - 5	0.50	(0.80)
G2	Yield	117	(65.36)	0 - 14	2.35	(3.09)
	Shift	167	(93.30)	0 - 15	4.21	(3.36)
	Total suggestibility	169	(94.41)	0 - 27	6.56	(5.77)
	Misinformation	46	(25.70)	0 - 4	0.34	(0.66)
ENTIRE GROUP	Yield	287	(77.15)	0 - 15	3.56	(3.71)
	Shift	357	(95.97)	0 - 18	5.65	(3.89)
	Total suggestibility	361	(97.04)	0 - 32	9.70	(7.10)
	Misinformation	117	(31.45)	0 - 5	0.42	(0.74)

It is evident that most children yielded to misleading questions and shifted their responses after the negative feedback was provided, while only a small number were susceptible to the post-event misinformation effect.

Interrogative suggestibility measures showed a normal distribution (skewness: Yield = 1.13, Shift = .74, Total Suggestibility = .63; kurtosis: Yield = .47, Shift = .02, Total Suggestibility = -.56), whereas post-event misinformation effect did not (skewness

= 2.31, kurtosis = 7.14). The post-event misinformation effect has been then dichotomised so to consider only the presence or the absence of the phenomenon.

To check for age effects on interrogative suggestibility a series of t-tests were conducted. A significant difference between the age groups was obtained for the interrogative suggestibility scores (Yield: $t_{(361.499)} = 6.43$, $p < .001$, $d = .66$; Shift: $t_{(370)} = 7.41$, $p < .001$, $d = .77$; Total Suggestibility: $t_{(365.341)} = 9.15$, $p < .001$, $d = .95$), indicating that overall younger children tended to be more suggestible than older ones (as shown in Table 4.4).

As a consequence of the overall results, the subsequent analyses of entire sample and two age-groups were conducted controlling for age. To investigate the relationships among the suggestibility measures and their associations with intelligence and executive functions, partial correlations were used. The correlation values are shown in Table 4.5. Interrogative suggestibility scores were positively correlated in most analyses. In addition, Yield, Shift and Total Suggestibility correlated negatively with Digit Span in the entire sample and in both age groups.

The Relationships between suggestibility and memory

The partial correlation values of the suggestibility scores and memory measures are indicated in Table 4.6 for the Entire Sample and the age-groups.

Table 4.5. Partial correlations among suggestibility and cognitive functioning measures.

		1.	2.	3.	4.
G1	1. Yield				
df = 184	2. Shift	.267***			
	3.Total Suggestibility	.567***	.820***		
	4. Misinformation	.111	.231***	.266***	
	5. Intelligence	.077	-.078	.026	-.126
	6. Working Memory	-.227**	-.165*	-.166*	-.001
	7. Attention	.170*	.009	.051	-.052
	8. Inhibitory Control	.168*	-.047	-.057	-.006
G2	1. Yield				
df = 175	2. Shift	.557***			
	3.Total Suggestibility	.865***	.899***		
	4. Misinformation	.198**	.128	.182*	
	5. Intelligence	.100	.063	.091	.138
	6. Working Memory	-.180*	-.227**	-.232**	.091
	7. Attention	-.034	-.029	-.036	.090
	8. Inhibitory Control	.043	.023	.037	.092
ENTIRE	1. Yield				
GROUP	2. Shift	.367***			
df = 362	3.Total Suggestibility	.647***	.849***		
	4. Misinformation	.149**	.187***	.225***	
	5. Intelligence	.088	-.014	.061	-.008
	6. Working Memory	-.176***	-.190***	-.186***	.049
	7. Attention	.089	-.010	.003	.014
	8. Inhibitory Control	.094	-.013	-.010	.028

*p<.05; **p<.01; ***p<.001

All suggestibility scores correlated negatively with free recall in the entire sample and the age-groups. Interestingly, the post-event misinformation effect was positively correlated with distortions but not with fabrications in the entire sample, whereas Shift and Total Suggestibility were correlated with fabrications but not with distortions in the entire sample and G1, while in G2 Shift was positively correlated with both fabrications and distortions. Yield was not significantly correlated with confabulations in the entire sample and G1, but was associated with distortions in G2.

Table 4.6. Partial correlations among memory and suggestibility measures.

		Yield	Shift	Total
G1	Free recall	-.175*	-.314***	-.382***
df = 184	Recognition	.091	-.318***	-.223**
	Distortions	-.117	.035	.013
	Fabrications	.090	.267***	.172*
G2	Free recall	-.316***	-.303***	-.350***
df = 175	Recognition	-.069	-.170*	-.139
	Distortions	.211**	.178*	.219**
	Fabrications	.028	.167*	.116
ENTIRE GROUP	Free recall	-.199***	-.301***	-.355***
df = 362	Recognition	.068	-.257***	-.190***
	Distortions	.043	.097	.092
	Fabrications	.077	.227***	.151**

*p<.05; **p<.01; ***p<.001

4.1.3. Discussion

Regarding the quantitative measures of memory, this study confirms the presence of pronounced age differences in the free recall task (Chae & Ceci, 2005; Chae, Goodman, Eisen, & Qin, 2011a; Roebers & Schneider, 2001b, 2002).

Furthermore, this study provides new evidence that distortions and fabrications should be considered to be two distinct phenomena (Gudjonsson & Clare, 1995). In fact, fabrications decreased with age, whereas distortions did not, the occurrence of fabrications was lower and the two measures were associated differently with cognitive functioning. However, the two phenomena were related: data showed that the children who tended to exhibit high levels of distortions committed also fabrications, and both confabulations were negatively associated with the free recall of the story.

Regarding the children's cognitive functioning the findings showed that it was only partially associated with quantitative measures of memory. Most previous studies have suggested that children with higher intellectual capacity recalled more aspects (Chae

& Ceci, 2005; Chae et al., 2011a; Eisen, Qin, Goodman, & Davis, 2002; McFarlane et al., 2002), with a few exceptions (Roebers & Schneider, 2001b).

Although theories suggest that executive functions play an important role in the coding, storage and retrieval of information from long-term memory (Alexander et al., 2002; Roebers & Schneider, 2005a), the findings of previous studies are inconsistent and contradictory. Working memory has been found to be related to recall in middle childhood, pre-adolescents and adults (Eisen et al., 2002; Lee, 2004) but not in preschoolers (Roebers & Schneider, 2005a). On the contrary, studies on other aspects such as inhibition, failed to detect correlations with free recall (Alexander et al., 2002; Melinder et al., 2006; Roebers & Schneider, 2005a; Scullin & Bonner, 2006). The findings of the present study are discrepant since a significant correlation was found with working memory only. These findings demonstrate that although executive functions play a role in memory, the different skills that underlie executive functions are not equally predictors of different memory abilities.

The interpretation of the findings concerning confabulations is more difficult, because most studies of the phenomenon have been conducted using recognition tests and experimental methods, while here a free recall task procedure was used. Generally, the former methods are considered more sensitive to the contents of memory and false alarms (that is "mistakenly accepting unpresented distractors" as words semantically related to the original stimulus of the list presented during the first phase of the experiment) provide the standard measure of this phenomenon (Brainerd & Mojardin, 1998). Generally studies have been conducted using the Deese/Roediger-Mc Dermott paradigm (DRM; Deese, 1959; Roediger & McDermott, 1995) but the comparison of these studies with the present findings would be questionable, because the issue of whether different procedures and experimental paradigms produce different forms of false memories, is as yet unresolved (Bernstein & Loftus, 2009; Otgaar & Candel, 2011). The present study was designed to investigate in depth confabulations in a semi-ecological context. The context was a

narrative event concerning a couple of children and their mother during a normal afternoon at home, in order to detect more ecological measures of episodic memory functioning. Nevertheless a few previous studies have investigated this complex phenomenon using recall tasks, which can be considered to be more ecological measures of confabulations of children than experimental procedures.

Regarding the inaccuracy of reported information under free-report conditions, age differences were found for fabrications and not for distortions. It is important to distinguish between different forms of confabulations, such as distortions and fabrications, because previous studies conducted using the Gudjonsson Suggestibility Scales (GSS; Gudjonsson, 1984, 1987) have supported the view that they may be considered two distinct and separate forms of memory errors, representing different psychological phenomena (Gudjonsson & Clare, 1995). The issue warrants further exploration. In fact, in the current study the distributions of these confabulations were not normal. Furthermore, fabrications occurred rarely: only 26% showed the phenomenon. It is possible that the two forms of confabulations are characterized by the large individual variability in middle childhood and that the sample of the present study might be too small to investigate this issue in depth. Therefore, generalizations about our results concerning these kind of spontaneous false memories (Brainerd & Howe, 1982) should be treated with caution.

Regarding the relationships between the different measures of memory, in the total sample analyses free recall was correlated negatively with distortions and fabrications. However, also the links between quantitative and qualitative scores of memory varied in the two groups. It is difficult to explain the differences in the links between these variables in the two age groups, also given their small sizes. Therefore, the issue warrants further investigation. Nevertheless previous studies have indicated that children, like adults with serious clinical disorders (in particular frontal lobe damage and psychiatric patients), tend to produce spontaneous false memories with greater frequency

when they have poor memory abilities (Gudjonsson & Clare, 1995; Gudjonsson & Young, 2010).

The interpretation of the relationships between spontaneous false memories in free recall and cognitive functioning are more complex because few studies have previously explored this issue. In the entire sample, confabulations did not tend to be correlated with cognitive measures. More specifically however, while distortions were not associated with cognitive measures, Fabrications were negatively related to working memory and positively related to inhibitory control scores. The results vary for the two age groups.

Studies of adults that have investigated the relations between confabulations and intelligence generally reported low or an absence of correlations (Gudjonsson & Clare, 1995; Gudjonsson & Sigurdsson, 1996). In contrast, previous research has indicated that inhibitory control plays a role in children's abilities to avoid false memories during recall (Alexander et al., 2002). To date, studies have not investigated working memory functioning and confabulations during free recall of stories, however using the DRM paradigm an association between working memory and the false recall of words was found in adults (Peters, Jelicic, Verbeek, & Merckelbach, 2007). Our finding that Fabrications are associated with working memory and inhibitory control scores, whereas Distortions are not, supports the view that they are two distinct phenomena (Gudjonsson & Clare, 1995). Finally, the positive association between inhibitory control and Fabrications in the Total Sample, and the positive association of Fabrication and working memory, even if the size of the effects is low, are surprising and warrant further exploration. They support the hypothesis that the individual variability concerning confabulations might be very wide, so our sample might be too small to detect the relationships of these phenomenon to cognitive functioning.

Regarding suggestibility, Study 5 supports data supports the view that children are more likely to yield to misleading questions, or to shift their responses, than to include misinformation in subsequent recall (Gobbo, 2000).

This study confirms the presence of pronounced age differences in interrogative suggestibility (see Bruck & Melnyk, 2004; Ceci & Bruck, 1993). Nevertheless, it is important to distinguish between the different interrogative suggestibility measures. Most studies have found that younger children give consistently more suggestible responses to misleading questions (Chae & Ceci, 2005; Chae et al., 2011a; Melinder et al., 2006). On the contrary, the literature concerning the age effect on Shift is controversial especially for preschoolers. In fact, some authors have hypothesized that Shift scores follow an inverted U-shape developmental trajectory peaking at around 5-6 years of age (Finnilä et al., 2003; Hünefeldt, Lucidi, Furia, & Rossi-Arnaud, 2008; Scullin & Ceci, 2001; Scullin, Kanaya, & Ceci, 2002). Studies with school-aged children have obtained discrepant results. Most have indicated that children usually do not present strong stability in their judgments or responses when adults repeat the questions (Otgaar & Candel, 2011; Roma et al., 2011), but they do not show a significant decrease in Shift answers after negative feedback, in relation to their age (Danielsdottir et al., 1993; Endres, 1997; Scullin et al., 2002). Study 5 data supports the view that also in middle childhood the tendency to provide Shift answers decreases with age, as found in a few previous studies (Warren et al., 1991; Young, Powell, & Dudgeon, 2003).

The Post-event misinformation effect was measured through a free recall procedure after exposing children to misleading questions. Typically, in studies of the misinformation effect children are repeatedly exposed to misinformation over a longer period (the first recognition task and some weakly sessions). The effect here found could therefore be due to the passage of time in combination with the repetition of suggestions. The present study indicated that repeated exposure to post-event misinformation effect within a single session is sufficient for some children to include post-event misinformation effect into subsequent recall. Furthermore, free recall tests have been employed infrequently to assess the post-event misinformation effect. Interestingly, a significant decrease in the phenomenon with age was found.

Since only a few children showed the post-event misinformation effect in the present and in previous studies, it may be necessary to investigate why some children are more prone than others to include false information in their memory (Lee, 2004). In Study 5, the post-event misinformation effect derived from the difficulty in source-monitoring between the information provided by the story and that provided by the interview. Unfortunately, however, the data obtained does not allow to make inferences about whether post-event suggestions interfere with recollection by rendering memories irretrievable as in the case of creating access competition, or whether the trace of the original event was altered because of interrogative suggestions. This was because the post-event misinformation effect was tested immediately after exposing the children to the misleading information during the interview.

From a theoretical point of view different types of memory suggestibility exist (Bruck & Melnyk, 2004). Studies on adults and children have provided strong evidence of the multidimensionality of the concept of suggestibility, but different forms have rarely been examined in the same study. Some studies of children have investigated interrogative suggestibility and false memories creation for events which never occurred (McFarlane & Powell, 2002; Miles, Powell, & Stokes, 2004; Scullin et al., 2002), but most have investigated only the two types of interrogative suggestibility (Gudjonsson, 1984, 1992). To date no other study has investigated the relationship between susceptibility to interrogative suggestions and the reporting of misleading details suggested during the free recall of an event in the same trial with children, at least to our knowledge. The relationship between the different forms depends on the degree to which suggestibility is a trait which is generalizable across different contexts, but more studies are required to clarify this issue (McFarlane & Powell, 2002). In terms of interrogative suggestions, most studies were conducted using standardized tests, such as the Gudjonsson Suggestibility Scales (Gudjonsson, 1984, 1992) and the Video Suggestibility Scale for Children (McFarlane & Powell, 2002; Scullin & Ceci, 2001) that measure Yield

and Shift scores separately. Generally, Yield and Shift have seemed to be quite independent or poorly correlated, both in adults and children (Gudjonsson, 1984; Lee, 2004; Quas, Wallin, Papini, Lench, & Scullin, 2005; Scullin & Bonner, 2006; Scullin & Ceci, 2001; Scullin et al., 2002; Warren et al., 1991; Young et al., 2003). Furthermore, factorial analyses of the GSS indicated that they depended on separate factors (Gudjonsson, 1984, 1992), but it is important to note that Gudjonsson used a component solution with a Varimax rotation. This procedure implies an orthogonal relationship between factors, thus the Yield and Shift components are necessarily uncorrelated (Gignac & Powell, 2009). Our analyses in the total sample and the age groups, showed positive correlations between Yield and Shift, so the issue warrants further investigation.

Interestingly, Study 5 data supports the view that children are more likely to assent to misleading questions when asked than to include the post-event misinformation effect into subsequent recall (Gobbo, 2000), however interrogative scores are related to the subsequent post-event misinformation effect in children's recall. This implies that when children accept misleading information provided during an interview, the memory traces can be partially replaced by the post-event misinformation effect. In this case, measures of interrogative suggestion are not simply measures of a transitory acceptance of suggestion during the interview, but are also indices of the malleability of memory. Finally, the high inter-correlations between the different forms of suggestibility suggest that children who are more prone to suggestion under some circumstances are also more prone to be suggestible in others.

Significant correlations were found between interrogative suggestibility scores and working memory only, whereas fluid intelligence and other tests of executive functions were not associated with suggestibility measures. Intelligence has been more frequently employed to try to explain children's individual differences in suggestibility in relatively homogeneous age-groups, but a review of the literature suggests discrepant results. Studies in early and middle-childhood found correlations between the ability to

resist suggestive questions and verbal intelligence (McFarlane et al., 2002; Roma et al., 2011; Singh & Gudjonsson, 1992) or non-verbal intelligence (Danielsdottir et al., 1993; Geddie et al., 2000; Roma et al., 2011), but most studies reported low or an absence of significant correlations (Bruck et al., 1997; Bruck & Melnyk, 2004; Chae & Ceci, 2005; Eisen et al., 2002; Roebers & Schneider, 2001b). On the contrary, studies that compared children with mental retardation to their age-matched typically-developing peers showed that they were more suggestible to misleading questions (Bruck & Melnyk, 2004; Gudjonsson & Henry, 2003; Young et al., 2003). An explanation of these discrepant results may be that only below-average intelligence correlates significantly with suggestibility (Bruck & Melnyk, 2004; Danielsdottir et al., 1993; Henry & Gudjonsson, 2003; McFarlane et al., 2002), even though Roebers and Schneider (2001) did not find differences between school-aged children with high and low IQs. Findings from Study 5 support the hypothesis that intelligence may not be a reliable predictor of suggestibility for children with normal IQs (Bruck & Melnyk, 2004).

Previous studies with preschoolers and/or school-aged children found that different aspects of executive functioning are negatively related to children's suggestibility. The findings concerning working memory capacity in children are discrepant, but it is important to underline that only a few studies have investigated this issue in depth (Bruck & Melnyk, 2004). Most studies failed to detect any relationship with the susceptibility to yield to misleading questions (Roebers & Schneider, 2005a). Other studies found significant correlations but only with certain specific measures. Lee (2004) for instance found a significant correlation between span and Shift but not between the Total Suggestibility and Yield, as measured with the GSS; Clarke-Stewart, Malloy, and Allhusen (2004) found that short term memory was related to suggestion in only one experienced event in six. Regarding the misinformation-effect, whereas studies with adults have found a relationship with working memory (Zhu et al., 2010), studies with children have not confirmed the phenomenon (Lee, 2004). Study 5 findings support the

view that working-memory and the post-event misinformation effect are not related in children. Regarding inhibitory control, several studies have suggested that children who are less able to inhibit responses tend to make more errors, when asked misleading questions (Alexander et al., 2002; Melinder et al., 2006; Roberts & Powell, 2005; Schaaf et al., 2008). Furthermore, a relationship between lack of behaviour control and suggestibility was also found using indirect measures of children's behaviour such as questionnaires filled in by parents (Clarke-Stewart et al., 2004), but most studies did not detect significant correlations (Bruck & Melnyk, 2004; Quas & Schaaf, 2002; Scullin & Bonner, 2006). The present findings do not support the prediction that children with less controlled executive functions are more suggestible, but it is important to remember that the scores in executive functioning tests were very high. Finally, Bruck and Melnyk (2004) argued that interrogative suggestibility is principally associated with temperamental factors, whereas the misinformation effect is associated with cognitive variables. However, no significant correlation was found between the post-event misinformation effect and cognitive functioning, which would support this view.

Despite being limited by the size of the two age groups, the results of Study 5 confirm a negative association between quantitative memories and the vulnerability to suggestibility. Intuitively it would seem that the strength of a memory for an event should be inversely related to the suggestibility for that same event. Several studies of adults and adolescents have supported this view (Gudjonsson & Singh, 1984; Gudjonsson & Young, 2010; Lee, 2004; Liebman et al., 2002). Further, children tend to be more suggestible when they receive only a single presentation of an event rather than repeated presentations (Clarke-Stewart et al., 2004) and elaborate verbal rehearsals of events may help to strengthen resistance to suggestive questions (Kulkofsky, 2010). However having a good memory does not necessarily prevent one from providing suggested responses, in fact, the relationship between memory and suggestibility has not been consistently observed in previous investigations with children (Chae & Ceci, 2005). Most studies that used a

standardised test for assessing suggestibility found a negative correlation between free recall of events and yielding to misleading questions (Danielsdottir et al., 1993; Finnilä et al., 2003; Gudjonsson, 1984; McFarlane et al., 2002; Scullin & Bonner, 2006; Warren et al., 1991), while only few indicated Shifts in responses after negative feedback (Danielsdottir et al., 1993; Warren et al., 1991). The finding that the two measures of suggestibility are often differentially related to memory provides evidence for the hypothesis that different types of suggestibility exist, even within the same paradigm (Gudjonsson, 1984; Liebman et al., 2002). By contrast, in the study by Roebers and Schneider (2002) most correlations between the children's ability to provide detailed recall of a video story and their ability to resist suggestions were not significant; (Quas et al., 2005) did not find any association between the VSSC recall and subsequent suggestibility measures. Furthermore, Kulkofsky (2010) found that free recall of an experienced event was not related to yielding to misleading questions. However, it is difficult to ascertain the degree to which these findings relate to memory ability per se or merely to poor recall for the story used. Other studies found associations between suggestibility and memory for unrelated events (Quas et al., 2005; Scullin et al., 2002). Overall, the findings of this study are consistent with the trace-theory which proposes that weaker memory traces are more vulnerable to suggestion than stronger traces.

Study 5 has important limitations and most concern the assessment of the cognitive functioning of children. First, to assess the intelligence of the children only one test was administered, but it is possible that different forms of intelligence are variously related to the different dimensions of suggestibility. Second, the scores of Go-No Go and TAU were quite high, hence it could be that the ceiling effects on these measures reduced the chances of finding significant relationships with memory and suggestibility. Third, a limitation related to our quantitative measure of memory. In fact, it is difficult to ascertain the degree to which our results relate to memory ability per se or merely to poor recall of the story that we used. Fourth, it is important to underline that most of the statistically

significant correlations tended to be quite low (.10 to .20), so the links among these variables might be considered to be weak. Fifth, the sample size of the children might not be considered sufficient to investigate the false memory phenomenon. Discrepant results concerning the total sample and the age group analyses show that the individual differences between the children were very great and only bigger samples might be able to represent the variability of the population in middle childhood.

Individual differences in children's suggestibility: The role of theory of mind, executive functioning and interviewer behaviour.

Theory of mind is the ability to recognise, in themselves and in others, the presence of mental states and to use them to understand and predict human behaviour (Premack & Woodruff, 1978). This ability develops rapidly and shows a dramatic conceptual change during preschool years (Wellman, Cross, & Watson, 2001), reaching a peak around four years; after this turning point, children usually understand that two people may have different representations of reality (Gopnik & Astington, 1988). However, the development of theory of mind continues during school-age years. When entering school, children usually succeed in the second order false belief, i.e. a kind of sophisticated recursive thinking in which a representation is included in another ("I think that you think that X thinks Z") (Perner & Wimmer, 1985), provided that they were able to understand first order false belief. From the theoretical point of view, there is evidence that first and second order false beliefs are headed by the same underlying skills, i.e. the understanding that beliefs can be false with respect to reality (Lecce, Cavallini, & Pagnin, 2010). Mentalization ability still develops during school-age years, when children become able to refer to mental states and explain the internal states that motivate behaviour. Recursive theory of mind ability continues to develop through the school years (Liddle & Nettle, 2006). The use of mental terms increases between seven and nine years (Meins, Fernyhough, Johnson, & Lidstone, 2006) and metacognition ability appears.

Children's ability to understand others' mental states develops along with improvements in cognitive abilities such as memory and executive functions, as well as with the enrichment in the vocabulary of mental states. Several theories have been

proposed to explain the relationship between theory of mind and executive functions (see Doherty, 2008). Inhibitory control and working memory are strongly related to the development of theory of mind (Carlson & Moses, 2001; Carlson, Moses, & Claxton, 2004). Executive functioning includes several skills needed for goal-directed behaviour such as inhibitory control, planning, strategy development, persistence, and working memory (Brocki & Bohlin, 2004). Working memory is a dynamic but limited capacity system which mentally maintains and stores information, while selectively processing new data to allow mental activities like reading, planning, reasoning and problem-solving (see Andrade, 2001; Conway, Jarrold, Kane, Miyake, & Towse, 2008).

The parallelism between the theory of mind development and the reduction of the vulnerability to interrogative suggestibility has led some authors to study this association.

In their review of the literature, Bruck & Melnyk (2004) considered eleven studies which investigated the relationship between theory of mind and suggestibility, finding inconsistent results. In the subsequent decade, other studies were conducted to deepen the understanding of this issue, but to date, there is no clear conclusion about it. Several studies confirmed the presence of a negative relationship between suggestibility and theory of mind (Bright-Paul, Jarrold, & Wright, 2008; Karpinski & Scullin, 2009; Melinder, Goodman, Eilertsen, & Magnussen, 2004; Parisi, Manno, & Fadda, 2007; Scullin & Bonner, 2006; Thomsen & Berntsen, 2005; Welch-Ross, 1999a, 1999b; Welch-Ross, Diecidue, & Miller, 1997), but there are studies that found no significant relationships (Quas & Schaaf, 2002; Roebers & Schneider, 2005a). Moreover, Templeton and Wilcox (2000) found that children who passed the false belief task showed a higher suggestibility than did children who failed the theory of mind task. They justified their finding as a support for the biased-guessing hypothesis (McCloskey & Zaragoza, 1985), that is, stating that children could keep in memory both accurate and inaccurate representations of the event, but as a consequence they were unsure which representation was accurate, guessing which answer was correct.

Theory of mind could be associated with either a reduced suggestibility or a heightened suggestibility. When children understand that a person may have a set of beliefs and understandings distinct from their own, they may reject the ideas of others because they know that the other may be wrong. Indeed, theory of mind ability allows children to understand that the interviewer may have a representation of the event that is different from their own - a false belief - but this raises the question how the child could react. It is likely that this knowledge would protect the child from the suggestion but the child could choose to comply with the other's false belief (Bruck & Melnyk, 2004). However, the issue is still open and deserves further attention. In addition, studies have mainly focused on pre-schoolers, without investigating the phenomenon at a later age, when theory of mind skills become increasingly sophisticated and multifaceted.

Children's theory of mind could largely affect the performance of children witnesses, because it underlies both children's understanding of the beliefs, intentions, and motivations of people involved in the event they need to disclose and their ability to understand that the interviewer may have incongruent beliefs from their own (Snow, Powell, & Sanger, 2012b). A child who must testify in a legal forum is in front of an interviewer that may ask leading questions and could have mistaken beliefs on the event, transmitting messages with his or her facial expression and the prosody of his or her phrases. Furthermore, the questions may contain terms referring to internal states, and the child may be able to understand them only if they have an adequate mastery of the psychological lexicon, the language including cognitive, emotional and volitional names and verbs referring to people's inner world (Bartsch & Wellman, 1995; Bretherton & Beeghly, 1982; Lecce & Pagnin, 2007).

Regarding executive functioning, previous studies showed that well-developed inhibitory control may contrast the tendency of children to accept misleading information (Alexander et al., 2002; Melinder et al., 2006; Roberts & Powell, 2005; Schaaf et al., 2008). Moreover, poor working memory could be a risk factor for heightening

interrogative suggestibility in children (Bottoms, Quas, & Davis, 2007; Lee, 2004; but see Roebers & Schneider, 2005) because it would make the questions comprehension more difficult by increasing the level of uncertainty. As shown in Study 5, working memory was negatively associated with interrogative suggestibility scores.

Theory of mind and executive functioning development are intertwined (Carlson, Mandell, & Williams, 2004; Hughes & Ensor, 2007) and it has been shown that a common process underlying executive functioning and theory of mind tasks accounted for variability in children's false reports (Poole, Dickinson, Brubacher, Liberty, & Kaake, 2014). The joint effect of these factors could also explain part of the variability in children's susceptibility to interrogative suggestibility.

5.1. STUDY 6. The relationship between interrogative suggestibility, theory of mind and executive functioning

Study 6 sought to provide an examination of the relation between continuous measures of theory of mind, executive functioning, and interrogative suggestibility in school-aged children. It would be useful both for theoretical and applied purposes to identify those factors responsible for individual differences in suggestibility. From an applied point of view, identifying factors involved in individual differences in suggestibility may help to predict the credibility of the testimony; from a theoretical point of view, it is interesting to find out to what extent cognitive and social mechanisms exert an effect on the individual predisposition to suggestibility.

The influence of theory of mind on interrogative suggestibility is still controversial and has not yet been studied in children of school age. In order to investigate different aspects of the phenomenon, the construct of the theory of mind was studied by using trials based on perceptual cues, such as a photograph of a person's eyes and the intonation of a human voice while pronouncing a sentence. Two separate components of

theory of mind were examined, namely the ability to infer mental states from acoustic cues and the ability to attribute a correct emotion by observing the facial expression to account for each child's experience of different sensory channels. The choice not to use tasks that rely on language was made to avoid the interfering effect of language ability implicated in more complex theory of mind tasks.

In line with most previous studies involving pre-schoolers (Bright-Paul et al., 2008; Karpinski & Scullin, 2009; Melinder et al., 2004; Scullin & Bonner, 2006; Thomsen & Berntsen, 2005; Welch-Ross, 1999a, 1999b; Welch-Ross et al., 1997), a negative relationship was assumed, hypothesizing that well-developed theory of mind abilities were associated with a lower vulnerability to suggestibility.

Moreover, executive functioning was assessed by using two different tasks aimed to evaluate two different components of this complex construct: working memory and inhibitory control. It was hypothesized that both these executive functioning components may be related to measures of interrogative suggestibility.

5.1.1. Method

Participants

This study took place at three primary schools in Northern Italy. The participants were 101 children (46 girls, 55 boys) ranging in age from 6.08 years to 9.15 years (M = 7.80, SD = .96)[23]. Informed written parental consent was obtained for all the children participating in the study. All participants were from middle-class families whose primary language was Italian.

[23] Martina Gueli and Lisa Dallera collected the data for the present study.

Instruments

The Bicocca Suggestibility Scales-B form, short version (BISS-B22), was used to investigate the children's memory and their vulnerability to interrogative suggestibility. The following measures were computed:

- *Free recall* = the amount of information correctly reported during the immediate recall of the story (maximum = 49).
- *Yield* = the number of leading questions to which the child has succumbed during the administration of the first series of questions (maximum = 14).
- *Shift* = the number of questions to which the child has changed his or her response after the negative feedback (maximum = 22).
- *Misinformation effect* = the number of details suggested during the questioning phase and subsequently reported in the delayed free recall of the story.

Voice Test of Mental States (Test Vocale sugli Stati Mentali - TVSM)

The Voice Test of Mental States (TVSM, Cavalli, Fabio, Sempio Liverta, & Marchetti, 2013) is a test used to investigate theory of mind development in school-aged children. It consists of 20 short segments of recorded audio dialogs to which the child listens and then chooses which one best describes the mental state of the speaker among 4 possibilities. Every recorded sentence is semantically neutral; therefore, the choice is based on the vocalization indices. The child also completes a control test. The entire protocol is considered valid if the child correctly identifies the gender of the speaker for at least 18 of 20 presented items. The test is equipped with a glossary that can be used whenever the child has difficulties in understanding a specific word. The raw score on the test can be calculated by giving one point for each correct answer (maximum = 20). This score can subsequently be transformed into a corresponding T-score specific for gender and age according to the test manual.

Eyes Test – Children Version (Test degli Occhi versione Bambini – TOB*)*

The Eyes Test (TOB, Castelli, 2010) is the Italian version of the Children's "Reading the Mind in the Eyes" Task (Baron-Cohen et al., 1997), a test used to investigate theory of mind development, which is operationalised as the inference of another person's mental state from photographs of the person's eyes. It consists of 28 photographs of a person's eyes, and the subject is asked to choose from 4 verbal labels the term that best describes how the depicted person feels. These labels are different for every photograph. The choice is between four mental state terms, some of which refers to basic emotions (e.g. sad, angry), and others of which are more complex, (e.g. confused, hopeful). There is also a control test requiring the participant to correctly indicate the gender of at least 9 of the photographed people to exclude randomness of the answers and to check if any deficits on the Eyes Task could be attributed to general deficits in face perception or perceptual discrimination (Baron-Cohen et al., 1997). A glossary is included and contains definitions and examples of the words on the answer sheet, which can be consulted whenever the subject has doubts about the meaning of some words. One point is given to each correct answer (maximum = 28). Because standardised scores were not available, raw scores were used.

Digit Span Test

The Digit Span Test (Wechsler, 1991) is a test used to measure the children's working memory. Strings of digits increasing in length are read aloud and, after every string, the child is asked to repeat it in the normal front to back order (forward). When the child fails in two consecutive trials having the same number of digits, new strings are read and the child is required to repeat in the back to front order (backward). The total number of correctly reproduced strings is considered a measure of verbal working memory. The children's raw scores of the cognitive tests were converted into the

normative scaled scores, according to the scoring procedures specified in the manuals (range: 1-19).

Go-No Go Test (Test delle ranette)

The Go-No Go Test is an auditory attention test drawn from the Italian battery for ADHD (Marzocchi et al., 2010). This test allows to evaluate the executive functioning of the child, regarding selective attention and inhibitory control. The child is asked to make a dot with a pen in a sequence of 12 cells every time a specific sound is presented and to refrain from doing so, when a different sound is presented. The child's raw score was converted into the normative percentiles scores, according to the scoring procedures specified in the manuals.

Procedure

The study was presented to the participants as a study about memory. Informed written parental consent was obtained for all the children participating in the study. Firstly, an individual meeting lasting about 30 minutes had been arranged for all the participants. During this meeting, all the children were administered the BISS-B22. During the "distracting" break expected between the immediate free recall and the questioning phase, the children completed the Go-No Go Test and the Digit Span Test. A further group meeting was organized: children in small groups of up to 8 – balanced for age - were administered the TOB and the TVSM.

5.1.2. Results

First, memory, suggestibility and theory of mind measures were examined. The descriptive statistics of all memory, suggestibility and theory of mind measures are shown in Table 5.1.

A closer look to the children's behaviour showed that only three children (3%) never endorsed the suggestion of misleading questions. Similarly, only three children never changed their previous answers after the negative feedback. In relation to the post-event misinformation effect, 15.8% of children showed this phenomenon. Most of them reported an incorrect information suggested during the questioning phase only once, one child twice and another child six times. This measure has been subsequently considered as dichotomy, by distinguishing children who did show the effect from those who did not.

Table 5.1. Descriptive statistics of memory, suggestibility, theory of mind, and executive functioning measures.

	Range	M	SD	Skewness	Kurtosis
Free recall	0 - 37	18.17	10.41	-0.50	-0.91
Yield	0 - 14	5.63	3.39	0.25	-0.71
Shift	0 - 20	5.43	4.11	1.35	2.36
Eyes Test	9 - 24	16.08	3.71	-0.14	-0.57
Voice Test	28 - 64	47.52	8.23	0.02	-0.59
Digit Span	3 - 15	8.23	2.99	0.30	-0.84
Go-No Go	5 - 95	54.29	29.17	-0.50	-1.09

To check whether there were gender differences in the way the children performed on memory and suggestibility tasks, a series of independent samples t-tests were conducted. None of the considered measures was significantly different between boys and girls (ps > .05). Moreover, a χ^2 test was conducted to check the effect of gender on the presence or absence of the misinformation effect, showing no significant differences ($\chi^2 = .34, p = .56$). Therefore, all the subsequent analyses were conducted on the collapsed data.

Pearson's bivariate correlations were then performed to investigate the effect of the age of the participants on the above-mentioned variables. Results showed statistically significant relationships with all the indices considered. Age was positively correlated

with free recall (r (93) = .71, p < .001). Alternatively, age was negatively correlated with Yield (r (101) = -.35, p < .001) and Shift (r (101) = -.35, p < .001).

An independent-samples Mann Whitney U test was conducted to investigate the effect of age on the presence of the misinformation effect. We did not find significant differences in child age for children who incurred in the post-event misinformation (M = 7.89, DS = .76) and those who did not incur in this phenomenon (M = 7.79, DS = .96), p = .95.

Because of the overall results, the subsequent analyses were conducted controlling for age.

The Pearson's partial correlation values between memory, suggestibility, and theory of mind are displayed in Table 5.2. Regarding the relationship between memory and suggestibility, the Yield score was negatively associated with the free recall performance. This means that a higher amount of information correctly recalled immediately after the reading of the story was related to a lesser tendency to be misled by suggestive questions.

The Yield score was significantly associated with the Shift score, i.e. the children who were more likely to succumb to misleading questions changed their answers after the negative feedback more often than the children who resisted the influence of suggestive questions.

Table 5.2. Pearson's partial correlations among memory, suggestibility, theory of mind, and executive functioning measures.

	1.	2.	3.	4.	5.	6.	7.
1. Free recall	1.00						
2. Yield	-.27*	1.00					
3. Shift	-.10	.37***	1.00				
4. Eyes Test	-.10	-.17	-.21*	1.00			
5. Voice Test	.10	-.25*	-.01	.22*	1.00		
6. Digit Span	.02	-.25*	-.18	.10	.00	1.00	
7. Go-No Go	.06	.01	.20*	.17	.19	-.09	1.00

* p < .05, ** p < .01, *** p < .001.

Regarding theory of mind, the children's performance on the Eyes Test was negatively associated with their tendency to Shift the original response after negative feedback. The higher the ability to detect the emotional state on a visual basis, the fewer the changes in the answers after the negative feedback.

The Yield score was negatively correlated with the performance on the Voice Test of Mental States. The higher the ability of the children to understand the emotional state on an auditory basis, the less often they were misled by suggestive questions.

To assess whether theory of mind skill contributed independent predictive power to a model controlling for age and free recall, a hierarchical regression analysis was conducted with the Yield score as the dependent variable. The results are displayed in Table 5.3.

Table 5.3. Hierarchical regression analyses predicting the Yield score with age, free recall score, theory of mind, and working memory.

		B	SE B	β
Step 1				
	Constant	15.43	2.78	
	Age	-1.26	0.35	-.35***
Step 2				
	Constant	10.66	3.27	
	Age	-0.38	0.49	-.11
	Free recall	-0.11	0.05	-.35*
Step 3				
	Constant	16.15	3.65	
	Age	-0.29	0.46	-.08
	Free recall	-0.10	0.04	-.31*
	Voice Test	-0.09	0.04	-.21*
	Digit Span	-0.27	0.11	-.24*

Note: R^2 = .13 for Step 1, ΔR^2 = .06 for Step 2 (p = .01), ΔR^2 = .10 for Step 3 (p = .004). Final R^2 = .29.
*p < .05, **p < .01, ***p < .001.

Age was entered in Step 1 and accounted for 13% of the variation in yield score. This model was statistically significant, F (1, 88) = 12.63, p = .001. Free recall was added at Step 2 and contributed to a statistically significant change, F (2, 87) = 10.01, p < .001

and the model accounted for additional 6% of the variation in yield score. The Voice Test of Mental States and Digit Span Test scores were added in Step 3 and contributed to a statistically significant change, F (4,85) = 8.52, $p < .001$ (final $R^2 = .29$). Both theory of mind ($sr^2 = .04$) and digit span ($sr^2 = .05$) emerged as independent predictors of children's tendency to Yield to misleading questions.

Then, it was of interest to deepen the role of theory of mind on the children's tendency to change their original answer after receiving negative feedback from the interviewer. Therefore, the different choices made by the children underlying the Shift score were explored. Particularly, the direction of the change for each response given by the children was considered. Hence, three possible trends were considered: the "negative shift", (M = .2.90, SD = 2.74, *Skewness* = 1.37, *Kurtosis* = 2.22), from an accurate answer to an incorrect one; the "positive shift" (M = .89, SD = 1.18, *Skewness* = 2.51, *Kurtosis* = 9.27) resulting in an increased accuracy during the repeated administration of the question; the "neutral shift" (M = 1.63, SD = 1.94, *Skewness* = 1.41, *Kurtosis* = 1.59), from a wrong answer to another equally incorrect (e.g. "Is the grandma's dog name Bobby or Toby?", the child chooses first one and then the other name, when actually in the story there are no dogs).

After, the relationship between theory of mind and executive functioning with the different components of the Shift score was explored by using Spearman's partial rank correlations. Non-parametric tests were used because of the peculiarity of these scores, the violation of normality and the high presence of 0. A significant relationship emerged between the negative shift and the Eyes Test score (Spearman's partial ρ = -.22, p = .03): the higher the ability to infer the mental state of other people by looking at their eyes, the less likely the child would succumb under the interrogative pressure by replacing an initial correct response with a wrong one. We found a positive correlation between the Go-No Go score and the positive shift score (Spearman's partial ρ = .26, p = .01), i.e. the children with greater selective attention abilities were more likely to use the negative feedback to

improve the accuracy of their responses than the children with less-developed attention skills. No other significant relationships were found (*p*s > .05).

Finally, a series of independent-samples Mann Whitney U-tests was conducted to assess whether there were differences in memory, interrogative suggestibility, theory of mind ability, and executive functioning between the children who showed the misinformation effect and the children who did not (see Table 5.4.). There were statistically significant differences for the Yield score: the children who showed the misinformation effect had a higher Yield score than the children who did not incorporate suggestive information in their delayed free recall. Furthermore, the children who exhibited the misinformation effect had a lower working memory skill than the children who did not. Surprisingly, the children who reported at least one suggested element in the delayed recall had a more developed inhibitory control ability than the children who never recalled post-event suggested information.

Table 5.4. Descriptive statistics of individual differences scores of children who showed and did not show the post-event misinformation effect.

	No misinformation effect			Misinformation effect			
	Range	*M*	*SD*	Range	*M*	*SD*	p
Free recall	0 - 37	18.49	10.52	0 - 27	16.63	10.05	.48
Yield	0 - 14	5.26	3.3	2 - 13	7.56	3.39	.02*
Shift	0 - 18	4.96	3.61	1 - 20	7.25	6.03	.22
TOB	9 - 23	16.11	3.53	9 - 24	15.6	4.66	.55
TVSM	28 - 64	47.2	8.39	43 - 59	50.75	5.09	.08
Digit span	3 - 15	8.53	3.01	4 - 12	6.88	2.73	.03*
Go-No go	5 - 90	51.73	28.95	20 - 95	69.67	20.22	.03*

5.1.3. Discussion

The present study sought to investigate the role of school-aged children's theory of mind and executive functioning in providing accurate recall, yielding to misleading questions, changing their answers in response to negative feedback and incorporating suggested information in their subsequent report.

First, we found support to the hypothesis that the strength of the memory trace is associated with memory accuracy and resistance to suggestion. However, memory for the story, assessed through the free recall task, was not related to children's working memory ability. Conversely, Yield score was negatively related to working memory. This result is consistent with previous studies (Caprin et al., 2016; Lee, 2004). The negative relationship between suggestibility and theory of mind is in line with most of the previous studies conducted with pre-schoolers (Bright-Paul et al., 2008; Karpinski & Scullin, 2009; Melinder et al., 2004; Scullin & Bonner, 2006; Thomsen & Berntsen, 2005; Welch-Ross, 1999a, 1999b; Welch-Ross et al., 1997). The ability to infer the other's mental states based on vocal cues made a significant contribution to variance in Yield score over and above child age and the memory of the story. On the contrary, the capacity to understand other's mind by looking in their eyes was related to the child's tendency to change an accurate response to one the child believed would please the interviewer. This lends some support to the hypothesis that a well-developed theory of mind is a protective factor against interrogative suggestibility. Therefore, a child with a poor theory of mind skill will find it hard to juggle different perspectives and several kinds of background information to provide structurally as well as qualitatively adequate narratives (Snow, Powell, & Sanger, 2012a). Moreover, the progressive development of the mentalization ability could partly explain the age differences observed in vulnerability to suggestibility.

The working memory skill and the ability to infer other's mental states seemed to be protective factors against the likelihood of falling into the misleading questions trap.

The observed model about yielding to suggestive questions could be explained by referring to the socio-cognitive conflict (Doise & Mugny, 1981). Indeed, since the child is required to provide an adequate answer during a social interaction, a well-developed theory of mind would detect the conflict between the child's memory and the misleading information suggested by the interviewer. The child then would assess the credibility of the contrasting information and choose the content of his or her memory, by resisting suggestibility. To do this, the child would need a sufficient memory of the event, as well as good working memory skill. Indeed, when answering an interviewer's questions, children should compare their own knowledge of a past event with the interviewer's knowledge of that same event. In the case of misleading questions, children would need to rely on their working memory resources to detect the contradiction between the interviewer's suggestions and their representation of reality.

It should be noted, however, that only the performance on the theory of mind task on an auditory basis correlated significantly with the Yield score. Theory of mind based on visual stimuli, i.e. the ability to infer the mental state by looking at the eyes of another person, was not associated with the answers given to suggestive questions. Considering that an interview largely relies on verbal information, it is not surprising that the ability to infer the content of another's mind from the intonation of the voice is a more reliable cue for suggestion than the facial expression. Since an early age, children are more likely to refer to audio rather than visual cues to infer the intentions of the adult and modify their behaviour in response to it (Mumme, Fernald, & Herrera, 1996; Vaish & Striano, 2004). The preference of audio stimuli in reading others' mind has been proved even in a study where police officers were involved in a lie and truth accuracy task: those who were exposed to audio cues only performed better than the participants who were exposed to visual information only (Mann, Vrij, Fisher, & Robinson, 2008).

The second component of interrogative suggestibility, i.e. the Shift score, was related to inhibitory control and to the performance on the Eyes Test. In addition, a deeper

analysis of the children's tendency to change their answers after negative feedback revealed that children with more developed executive functioning were more able to take advantage of the interviewer's feedback to improve the accuracy of their answers. We found inhibitory control to have a positive role in predicting the tendency to correct the previous answers. It is possible that children with weaker inhibitory control have difficulties in keeping track of the original elements while suppressing the predominant response suggested by the interviewer (Alexander et al., 2002; Bruck & Melnyk, 2004). Therefore, they would be unable to replace their wrong answer with the right one when given a second chance to answer the questions.

Furthermore, the data supported the view that children are more prone to yield to suggestive questions or to change their answers as result of the interviewer's negative feedback, than to include misinformation in the subsequent recall. This is consistent with Gobbo's (2000) study involving pre-schoolers and school-aged children involved in a staged event.

In relation to the post-event misinformation effect, the present study indicated that the exposure to suggestive information within a single session is sufficient for some children to report some of the suggested elements in their subsequent recall as if they were originally present in the story. In contrast to our expectations and to the results of a previous study with the Bicocca Suggestibility Scales (Caprin et al., 2016; see Chapter 4), we did not find a significant reduction in the exhibition of this phenomenon in older children compared to the younger ones. However, only 16 children showed the misinformation effect. Therefore, it is possible that the analysis did not have enough statistical power to detect the effect of age.

The children who reported the misleading information in their delayed free recall had also a higher tendency to yield to suggestive questions and a poorer working memory skill than the children who did not show the misinformation effect. However, it is surprisingly that children who exhibited the misinformation effect had a better-developed

inhibitory control than children who did not. By looking at all the results it seems that, even if children are more likely to yield to suggestive questions in response to a specific question than to include the post-event misinformation effect into their subsequent recall (Gobbo, 2000), they can provide erroneous information when answering an open-ended question if previously exposed to suggestion. A weaker working memory could make it difficult to monitor the original source of information, and having assented to the interviewer's suggestions would heighten the risk of believing in it. It may be that, once the child has accepted a misleading element as true, the original memory would become the "interfering" response to suppress. We cannot draw a clear conclusion from our results, and future research should deepen understanding of why inhibitory control is associated with both an increase of positive Shift and an augmented tendency to exhibit the misinformation effect.

5.2. STUDY 7. The role of theory of mind and interviewer behaviour on interrogative suggestibility.

One essential requirement for interrogative suggestibility to occur is the social interaction between the interviewee and the interviewer (Gudjonsson & Clark, 1986). Therefore, to achieve a better understanding of this phenomenon, both individuals involved must be considered. Studies that have investigated interviewer behaviour during interviews revealed that adults' memory accuracy is affected by a friendly or an abrupt attitude (Bain & Baxter, 2000; Bain, Baxter, & Fellowes, 2004; Baxter & Boon, 2000; Baxter, Jackson, & Bain, 2003). However, only one study showed that the interviewer's demeanour influenced adults' tendency to succumb to misleading questions (Bain et al., 2004). The willingness to change previous answers after receiving negative feedback instead seems to be more sensitive to the social aspects of suggestibility (Bain & Baxter, 2000; Baxter & Boon, 2000). Children's performance during an interview is influenced

by the demeanour of the interviewer much more than adults' performance is. In particular, social support exerts a key function for the accuracy of children's reports (Carter, Bottoms, & Levine, 1996; Davis & Bottoms, 2002; Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; Quas et al., 2005) and for the disclosure of abuse during the interview (Lewy, Cyr, & Dion, 2015). Laboratory studies showed that when children are questioned by a supportive rather than a non-supportive interviewer, suggestibility is reduced (see Bottoms, Quas, & Davis, 2007 for a review). Moreover, interviewer social support was found to have a positive effect on children's informativeness during substantive questioning in the naturalistic context of real investigative interviews (Teoh & Lamb, 2013).

Although a positive role of social support on children's reports has been demonstrated, it is of theoretical and practical interest to understand the precise conditions under which social support improves the accuracy of children's testimonies. Individual differences may interact with the effects of social support on memory and suggestibility by rendering some children more vulnerable to intimidating behaviour from the adult who is questioning them. In addition, it is useful to identify children who most need to be socially supported during interviews.

After finding support for the hypothesis about the protective role of individual differences in theory of mind ability on children witness skills, Study 7 aimed to deepen understanding of this relationship in the interplay with external factors that could occur during a formal interview. Therefore, the interaction between theory of mind and the interviewer's behaviour was investigated. It was hypothesized that all the children would be more accurate and less suggestible when interviewed by a friendly interviewer rather than an abrupt interviewer. Moreover, it was expected that the children with a well-developed theory of mind would better resist the negative effect of being interviewed by an abrupt interviewer than the children with less-developed theory of mind. It was assumed that the ability to infer the mental state of others may be a protective factor such

that a hostile attitude may not affect the child's performance; instead, performance may be affected if the child had poor theory of mind.

A second aim of the present study was to assess the role of suggestibility as a trait-like characteristic by examining the relationship between the BISS-A32 Yield scores measured in Study 2 and the children's acceptance of misleading questions during a structured interview about a video. It was hypothesized that the BISS Yield scores would predict the accuracy in the structured interview over and above the effects of children's age and theory of mind.

Furthermore, it was expected that the most suggestible children (according to the BISS scores) would be more vulnerable when interviewed by a non-supportive person than by a supportive one.

5.2.1. Method

Participants

The participants were 142 children (63 girls, 79 boys) ranging in age from 5.86 years to 8.88 years (*M* age = 7.36, *SD* = .95). They were sampled from four primary schools in Northern Italy[24]. Parental informed consent was obtained for each child prior to participation in the project. All the children who received parental written consent assented to participate in this study. Family demographic forms, including information on education and occupation of parents, were returned by 137 children (96.5%). Of those who responded, the children's families were predominantly of Italian ethnic heritage (78.2%) with low (21.8%), middle (43.7%) and high (31%) socioeconomic status according to information about parental education and occupation.

[24] Elisa Perosi and Laura Balduchelli collected the data for the present study.

Instruments

Bicocca Suggestibility Scales (BISS-A32)

The Bicocca Suggestibility Scales-A Form, long version (BISS-A32), was used to investigate the children's memory and their vulnerability to interrogative suggestibility.

The following measures were considered:

– *Free recall* = the amount of information correctly reported during the immediate recall of the story (maximum = 49).

– *Yield* = the number of misleading questions to which the child has given the wrong answer during the administration of the first series of questions (maximum = 16).

– *Shift* = the number of questions to which the child has changed his or her response after the interviewer's negative feedback (maximum = 32).

Voice Test of Mental States (Test Vocale sugli Stati Mentali - TVSM)

The Voice Test of Mental States (TVSM, Cavalli, Fabio, Sempio Liverta, & Marchetti, 2013) is a test used to investigate theory of mind development in school-aged children. It consists of 20 short segments of recorded audio dialogs to which the child listens and then chooses which one best describes the mental state of the speaker among 4 possibilities. Every recorded sentence is semantically neutral; therefore, the choice is based on the vocalization indices. The child also completes a control test. The entire protocol is considered valid if the child correctly identifies the gender of the speaker for at least 18 of 20 presented items. The test is equipped with a glossary that can be used whenever the child has difficulties in understanding a specific word. The raw score on the test can be calculated by giving one point for each correct answer (maximum = 20). This score can subsequently be transformed into a corresponding T-score specific for gender and age according to the test manual.

Eyes Test – Children Version (Test degli Occhi versione Bambini – TOB*)*

The Eyes Test (TOB, Castelli, 2010) is the Italian version of the Children's "Reading the Mind in the Eyes" Task (Baron-Cohen et al., 1997), a test used to investigate theory of mind development, which is operationalised as the inference of another person's mental state from photographs of the person's eyes. It consists of 28 photographs of a person's eyes, and the subject is asked to choose from 4 verbal labels the term that best describes how the depicted person feels. These labels are different for every photograph. The choice is between four mental state terms, some of which refers to basic emotions (e.g. sad, angry), and others of which are more complex, (e.g. confused, hopeful). There is also a control test requiring the participant to correctly indicate the gender of at least 9 of the photographed people to exclude randomness of the answers and to check if any deficits on the Eyes Task could be attributed to general deficits in face perception or perceptual discrimination (Baron-Cohen et al., 1997). A glossary is included and contains definitions and examples of the words on the answer sheet, which can be consulted whenever the subject has doubts about the meaning of some words. One point is given to each correct answer (maximum = 28). Because standardised scores were not available, raw scores were used.

Design and procedure

The study was approved by the Ethical Committee of the University of Milano-Bicocca. Informed written parental consent was obtained for all the children participating in the study. An individual meeting lasting approximately 40 minutes was arranged for all the participants. During this meeting, all children were administered the BISS-A32 and the *Raven's Coloured Progressive Matrices* (CPM, Raven, 1998) to assure that all the participants' IQs were above the 25th percentile. A further group meeting was organized, and all the children were administered the TVSM and the TOB collectively in

groups of about 8 members. The TVSM was administered immediately before the TOB for all children.

Study 7 comprised three phases and had a 2x2 factorial, between-participants design. The independent variables were individual theory of mind, “high” or “low”, and the experimental manipulation of the interviewer’s behaviour, “abrupt” or “friendly”.

Phase 1 – The BISS-A32 was administered to all the children.

Phase 2 – The children completed the two theory of mind tasks. The TVSM T-scores were used to create groups of children with high or low theory of mind. Since the previous study found that only the TVSM T-scores, not the TOB scores, significantly predicted Yield scores, the TVSM T-scores were used to create the different experimental groups. The children were allocated to two groups based on their mentalization ability (i.e., their T-score in the TVSM task). The median (*Median* = 52) was chosen as the dividing point: children whose scores were less than or equal to the median value were included in the low theory of mind group, whereas the others were placed in the high theory of mind group.

Phase 3 – the children of both groups watched a video and were then randomly assigned to one of the two experimental conditions. In the first condition, they were interviewed by an "abrupt" interviewer, and in the second condition, they were interviewed by a "friendly" interviewer.

To evaluate interrogative suggestibility, a brief video was created along with a specific interview protocol. The video was created as a cartoon to be more accessible to an audience of school-aged children. The video presents some ambiguous information while a narrating voice describes a robbery perpetrated against a lady.

Children were randomly allocated to one of two conditions: the interview with the abrupt interviewer or with the friendly interviewer. To avoid interviewer bias and to prevent the examiner's beliefs from influencing the child’s attitude during the interview, the experimenters were naïve to the content of the video shown. The two experimenters

did not watch the video and were blind to the children's inclusion in the high or low theory of mind group.

Each child watched the video individually on a computer monitor with headphones. Afterwards, the experimenter explained to the child that she had never seen the video and asked the child to tell her everything he could remember about the video with the purpose of consolidating the memory trace of the event.

The friendly interviewer smiled at the child and gave him or her indiscriminate positive feedback in a warm and encouraging tone ("Bravo! Very good job!"). Conversely, the abrupt interviewer gave indiscriminate mild negative feedback with a firm tone ("I was expecting something more. Oh well, let's go on").

Thereafter, there was a break of 20 minutes while the examiner proposed unrelated filler tasks to the child. After this break, both the interviewers gave the child pre-interview instructions by providing two ground rules: a statement about the interviewer's naiveté and an encouragement to say "I don't know" when the child did not know the answer ("Now I'm going to ask you some questions about the video. I've never seen the video, so I do not know what the correct answer is. If there are any questions about something you don't remember, do not to try to guess, but tell me you don't know"). Then, the friendly interviewer smiled at the child and added reassurance ("Anyway, don't worry. If you make some mistakes, nothing will happen to you"). Conversely, the abrupt interviewer maintained a stern face and said, "I recommend that you do not make mistakes!"

The interview was the same for both conditions and contained 38 questions. Ten were closed questions that focused on specific information but limited the response to yes or no (6 questions, for which 3 required "Yes" and 3 "No" as the correct answer) or asked the child to choose between two options (4 questions). These questions were about details actually present in the video because the children should have known the correct answer to those questions. There were 16 suggestive questions; half of these were yes/no

questions and the other half were alternative-choice questions that referred to details that were not presented in the video. Moreover, 9 questions were repeated immediately after the same question was presented.

The abrupt interviewer maintained a serious attitude at all times, never smiled and did not provide support to the child. However, at the end of the meeting, she smiled at the child and reassured him or her with positive feedback about overall performance to compensate for the negative feelings induced by the procedure. The friendly interviewer smiled often at the child and maintained supportive behaviour all the time.

The following indices were computed: Yield was the sum of the incorrect answers to the misleading questions. Moreover, the number of times the child provided an "I don't know" answer was considered for both the control questions and the suggestive questions.

The presentation of the video and the interview occurred in the classroom of the school individually for each child. Distracting tasks were administered over an interval of approximately 20 minutes between the immediate recall and the interview.

5.2.2. Results

Preliminary analyses revealed that the BISS Yield scores did not differ in terms of gender ($t_{(139.482)} = .89$, $p = .37$), so all the data were collapsed across this variable for further analyses.

Then, the children's scores on the BISS memory and suggestibility measures were considered: free recall ($M = 24.20$, $SD = 8.46$), Yield ($M = 7.15$, $SD = 4.78$), and Shift ($M = 7.70$, $SD = 4.05$). Pearson's correlations were calculated and revealed a significant positive association between child age and both memory scores (free recall: $r = .49$, $p < .001$). In contrast, correlations between age and the BISS suggestibility scores showed significant negative associations between age and Yield ($r = -.52$, $p < .001$) and Shift score ($r = -.46$, $p < .001$). In other words, the older children recalled more correct

information about the story and were less likely to succumb to leading questions than the younger children were. Furthermore, the older children changed their responses when told that they had made mistakes less often than the younger children did.

Partial correlations, controlling for age, showed a significant negative association between free recall and Yield score (r = -.30, p < .001). Conversely, there was no significant association between free recall and Shift score (r = -.12, p = .17). The more correct information the children could recount about the video during free recall, the less likely they were to be misled by the leading questions. Nevertheless, the number of correct responses to specific non-leading questions was not related to the child's tendency to succumb to suggestions. Regarding the responses that children gave after receiving mild negative feedback, the more correct answers the children gave to specific non-leading questions, the less likely they were to change their original responses. However, the number of changes was not associated with the amount of correct free-recall information.

Next, the scores obtained by the children in the control test for both the TOB and TVSM were checked. The performance of two children on the control test of the TOB was under the cut-off score specified in the test manual; similarly, five children scored below the cut-off point on the control test of the TVSM. Hence, the data from these seven children were excluded from the analyses.

The children's scores on the theory of mind tasks were considered: TOB (M = 15.44, SD = 4.31) and TVSM (M = 50.79, SD = 9.27). Partial correlations, controlling for age, were conducted. Both the theory of mind measures were significantly and positively associated with the BISS free recall (TOB: r = .17, p = .048; TVSM: r = .20, p = .02). The results showed a significant negative correlation between the TVSM T-scores and the BISS Yield scores (r = -.28, p = .001). However, the TOB scores were not significantly correlated with the BISS Yield score (r = -.13, p = .15), and the BISS Shift scores were

not related to any of the theory of mind tasks (TOB: r = -.08, p = .36; TVSM: r = -.12, p = .18).

To assess whether theory of mind contributed independent predictive power to a model comprising age and free recall score, a hierarchical regression analysis was conducted with Yield score as the dependent variable. Therefore, age was entered in Step 1. This model was statistically significant, F (1, 133) = 44.32, p < .001, and accounted for 25% of the variation in yield score. Free recall was entered in Step 2 and contributed to a statistically significant change, F (2, 132) = 30.33, p < .001, by accounting for additional 7% of the variation in yield score. Finally, TVSM T-scores were added in Step 3 and contributed to a statistically significant change, F (3, 131) = 23.83, p < .001 (final R^2 = .35). The results are displayed in Table 5.5. Theory of mind (sr^2 = .04) emerged as independent predictors of children's tendency to Yield to misleading questions, over and above child age and free recall.

Then, the children were allocated to two groups based on their mentalization ability (i.e., their T-score in the TVSM task). The median (*Median* = 52) was chosen as the dividing point: children whose scores were less than or equal to the median value were included in the low theory of mind group, whereas the others were placed in the high theory of mind group. A preliminary analysis was performed to exclude the possible interfering effect of pre-existing differences between groups. Approximately the same number of boys and girls were interviewed by the abrupt (40 males, 28 females) and the friendly (35 males, 32 females) interviewer (χ^2 (1, N = 135) = .59, p = .44). A series of t-tests were conducted, and no significant differences were found in the age of the children between the two experimental conditions (abrupt: M = 7.54, SD = 0.87; friendly: M = 7.27, SD = 0.98). No significant differences were found for free recall (abrupt: M = 24.20, SD = 9.02; friendly: M = 25.04, SD = 7.45) or Yield scores (abrupt: M = 7.00, SD = 4.63; friendly: M = 6.85, SD = 4.94) previously measured by administering the BISS-A32 (all ps > .05).

Table 5.5. Hierarchical regression analyses predicting the Yield score with age, free recall score, and theory of mind.

		B	*SE B*	*β*
Step 1				
	Constant	25.86	2.87	
	Age	-2.56	0.38	-.50***
Step 2				
	Constant	24.72	2.77	
	Age	-1.85	0.42	-.36***
	Free Recall	-0.17	0.05	-.29***
Step 3				
	Constant	29.10	3.13	
	Age	-1.81	0.41	-.35***
	Free Recall	-0.14	0.05	-.25**
	TVSM	-0.10	0.04	-.20**

Note: R^2 = .25 for Step 1, ΔR^2 = .07 for Step 2 (p = .001), ΔR^2 = .04 for Step 3 (p = .006).
*p < .05, **p < .01, ***p < .001.

The data obtained from the interview about the video were subjected to a 2 (ToM: low vs. high) x 2 (the interviewer's style: abrupt vs. friendly) analysis of covariance (ANCOVA) with age added as the covariate. Table 5.6 shows the children's answers in the interview about the video grouped into the different experimental conditions.

First, the number of correct responses to the non-leading specific questions about the video was analysed. The experimental condition ($F_{(1, 130)} = 0.84$, $p = .36$, $\eta^2 = .006$) and theory of mind ($F_{(1, 130)} = 3.30$, $p = .07$, $\eta^2 = .025$) did not have a statistically significant effect on the children's ability to provide correct responses to closed non-leading questions. The interaction was not significant ($F_{(1, 130)} = 1.37$, $p = .24$, $\eta^2 = .010$).

Table 5.6. Descriptive statistics of the children's answers about the video.

	Low ToM			High ToM		
	N	*M*	*SD*	N	*M*	*SD*
Correct answers to control questions						
Abrupt Interviewer	37	8.14	1.40	31	8.87	1.09
Friendly Interviewer	38	8.11	1.29	29	8.28	1.51
Yield to misleading questions						
Abrupt Interviewer	37	10.97	4.22	31	9.32	4.01
Friendly Interviewer	38	9.26	4.00	29	7.41	3.67

Then, the number of times the children were misled by suggestive questions about the video was considered. As expected, the results showed a main effect of theory of mind on Yield scores ($F_{(1, 130)} = 5.81$, $p = .02$, $\eta^2 = .043$). That is, children in the high ToM group were overall less suggestible ($M = 8.40$, $SD = 3.94$) than children in the low ToM group ($M = 10.11$, $SD = 4.17$). A main effect of the interviewer's behaviour on Yield scores ($F_{(1, 130)} = 13.33$, $p < .001$, $\eta^2 = .093$) showed that the children interviewed by the abrupt interviewer ($M = 10.22$, $SD = 4.18$) were more likely to succumb to misleading questions than the children interviewed by the friendly interviewer ($M = 8.46$, $SD = 3.94$). The interaction between theory of mind and the interviewer's behaviour was not statistically significant ($F_{(1, 130)} = .11$, $p = .75$, $\eta^2 = .001$). Although this interaction was not significant, it is remarkable that the highest number of incorrect answers to misleading questions was found for children with a poorly developed theory of mind who were interviewed by the abrupt interviewer. Conversely, the Yield score of children with a well-developed theory of mind who were interviewed in a friendly manner was the lowest.

Finally, the issues of the role of suggestibility as a trait-like characteristic and its interaction with the interviewer's behaviour were investigated. Pearson bivariate correlations were calculated between the BISS Yield scores and the incorrect answers to

misleading questions about the video. The results showed a strong relation between these measures ($r = .66$, $p < .001$). With age controlled, Pearson's partial correlations were still highly significant ($r = .58$, $p < .001$). We attempted to separate the individual contribution of the level of suggestibility (i.e., the Yield score previously obtained by the administration of the BISS-A32) from those of the theory of mind and the interviewer's demeanour in predicting responses to suggestive questions about the video. Therefore, the independent contribution of age, BISS Yield score, theory of mind, the interviewer's behaviour, and the interaction between BISS Yield scores and the interviewer's behaviour was considered. A hierarchical regression analysis was performed and the results are displayed in Table 5.7. Child age was entered in Step 1 and this model was statistically significant, $F(1, 133) = 23.30$, $p < .001$, and accounted for 15% of the variation in answers to misleading questions about the video. Then, the BISS Yield score, interviewer's behaviour and theory of mind score in Step 2. This model contributed to a statistically significant change, $F(4, 130) = 30.98$, $p < .001$, by accounting for additional 34% of the variation in yield score. Finally, the interaction between the Yield score and the interviewer's behaviour was added in Step 3. This model contributed to a statistically significant change, $F(5, 129) = 25.39$, $p < .001$, by accounting for additional 1% of the variation in yielding to misleading questions about the video (final $R^2 = .50$). Suggestibility and the interviewer's behaviour were significant predictors of the tendency to Yield to suggestive questions about the video. The model explained approximately 51% of the variability in the vulnerability to be misled by suggestive questions. The interaction between the BISS Yield scores and the interviewer's behaviour was excluded from the model since it did not significantly contribute explanatory power to the model ($p = .16$).

Table 5.7. Hierarchical regression analyses predicting the incorrect answers to the misleading questions with age, BISS Yield score, the interviewer's behaviour and theory of mind.

		B	*SE B*	*β*
Step 1	Constant	22.05	2.65	
	Age	-1.71	0.36	-.39***
Step 2				
	Age	-0.55	0.33	-.12
	Yield (BISS)	0.50	0.07	.58***
	Interviewer behaviour	-1.81	0.53	-.22***
	TVSM	-0.03	0.03	-.06
Step 3				
	Age	-0.56	0.33	-.12
	Yield (BISS)	0.58	0.09	.67***
	Interviewer behaviour	-1.85	0.53	-.22***
	TVSM	-0.89	0.03	-.07
	Interviewer behaviour x Yield	-0.15	0.11	-.13

Note: R^2 = .15 for Step 1, ΔR^2 = .34 for Step 2 ($p < .001$), ΔR^2 = .01 for Step 3 (p = .16). Final R^2 = .50.

*$p < .05$, **$p < .01$, ***$p < .001$.

5.2.3. Discussion

The results of the Study 7 provide evidence to support the hypothesis that theory of mind is a protective factor against vulnerability to interrogative suggestibility. First, a significant negative relationship was found between suggestibility and theory of mind measured with two different tests that operationalized the construct as understanding another's mind based on perceptual indices, such as the sound of the human voice and the gaze. However, this conclusion is valid only for the scores from the TVSM. This result means that the ability to infer the other's mental states from the intonation of his or her

voice would help the child to resist suggestions. When controlling for age and memory of the event, the contribution of theory of mind was still significant. Conversely, the theory of mind measure based on visual indices was not associated with Yield scores.

Then, children with a well-developed theory of mind resisted suggestions more than children in the low theory of mind group did. In line with most previous studies, which investigated this ability in preschoolers, children with high theory of mind showed a lower tendency to Yield to the interviewer's misleading questions than children with low theory of mind did (Bright-Paul et al., 2008; Karpinski & Scullin, 2009). This finding is in line with Bruck and Melnyk's (2004) hypothesis that theory of mind could account for individual differences in interrogative suggestibility, even in school-aged children.

The vulnerability to leading questions appeared to be biased significantly by the interviewer's style. Consistent with the initial hypothesis, the results showed that when the interviewer's attitude was gruff, children tended to be particularly vulnerable to suggestion. More specifically, they tended to Yield to misleading questions about the video more than children in the supportive condition did. This result confirms the findings of previous studies (Carter et al., 1996; Davis & Bottoms, 2002; Goodman et al., 1991; Quas et al., 2005). However, vulnerability to suggestibility was influenced by the adult's attitude in an apparently counter-intuitive way. The non-supportive person was firmer than the friendly interviewer in recommending that the children be as accurate as possible when answering the questions. Therefore, the children should have tried to be more accurate to comply with her, when in fact, the opposite occurred. We may assume that the firm demeanour of the adult may have caused to the children to have increased anxiety, which resulted in a lack of self-regulation by inducing worse performance than children in the other condition. This effect could be explained by hypothesizing that the abrupt behaviour of the interviewer may lead the child to experience intense anxiety that produces an increased feeling of uncertainty about the child's memories. This uncertainty

may result in a greater tendency to Yield to suggestive questions (Gudjonsson & Clark, 1986).

The third aim of the present study was to assess trait-like suggestibility (measured with the BISS-A32) in relation to the accuracy of children's responses to a structured interview about a video. These two suggestibility measures from different tasks were strongly related, and this relation had a large effect size (Cohen, 1988), even after age was partialled out. The regression model showed that the BISS-A32 Yield measure was the most important predictor of the accuracy of children's responses in the interview about the video, over and above age. This result provides support for the hypothesis that suggestibility is a stable individual trait and that BISS-A32 Yield scores are good predictors of children's performance on a similar task, even more than chronological age. This finding opens the possibility of introducing the Bicocca Suggestibility Scales as a forensic tool in the assessment of child witnesses. However, it is premature to determine the actual utility and predictive validity of this instrument. The results of the current study should be replicated in future research by introducing a comparison with suggestibility for a staged event.

This study has some limitations. The main limitation is the scarce ecological validity. Even though the findings can be considered for basic research, their value for applied contexts should be demonstrated with future studies in real interviewing contexts. Moreover, it is not possible to generalise these conclusions to children exposed to child abuse and neglect, who often show deficits in theory of mind (Cicchetti, Rogosch, Maughan, Toth, & Bruce, 2003; Pears & Fisher, 2005) and are more vulnerable to suggestion (e.g. Benedan et al., 2018; Curci et al., 2016). However, it seems plausible that theory of mind could be responsible for part of these children's well-known tendency to Yield to misleading questions. Future research is needed to fully understand the role of theory of mind in explaining individual differences in child witness performance.

The results of this study may improve the theoretical knowledge on the topic, which in turn may affect practice in the legal field. Indeed, the emphasis on individual differences in child witnesses is of both theoretical and practical value. The results shed light on some aspects underlying children's tendency to succumb to interrogative suggestibility. It would be important to assess the child's ability to be a witness by considering his or her level of suggestibility and the ability to interpret his or her own and others' mental states. The introduction of theory of mind tasks in the evaluation of the child's cognitive functioning would provide an overview of his or her ability to understand other's behaviour. This process would also allow professionals to be aware of each child's peculiarities and needs.

Finally, there are implications for people who interview children in investigative interviewing and forensics areas. Children are particularly sensitive to the interviewer's attitude; therefore, care must be taken. It may not be sufficient to avoid leading questions and negative feedback; it is also necessary to be aware of the intonation given to sentences and to the overall non-verbal behaviour.

Individual differences in children's suggestibility: the role of temperament.

A serious problem in the field of legal testimony is the knowledge that under some conditions everyone may succumb to memory failure or suggestion during an interrogation, and that this susceptibility to situational factors varies according to individual characteristics, which can moderate it or increase it (Gudjonsson, 1983). This phenomenon is especially well documented with children because, despite age-related differences, under the same conditions they vary considerably in their memory performances and susceptibility to interrogative suggestion: some provide very accurate recall of past experiences and are resistant to suggestions, whereas others provide little information and are highly prone to suggestion (Chae & Ceci, 2005). Several studies have explored the individual characteristics that may explain this variability, documenting a wide range of cognitive and temperamental factors (for a review see Bruck & Melnyk, 2004). It is obvious that in a social context such as a conversation or in an interrogation, individual cognitive and temperamental factors are intertwined and influence vulnerability to memory recall and suggestion (Roebers & Schneider, 2005a). However, to date most studies have been directed towards the investigation of cognitive factors, neglecting temperamental factors, despite the fact that the latter may consistently affect the recall of an event as well as mediating the ability to resist suggestive questions (Raju, Qin, Schaaf, & Goodman, 1997).

Previous studies with adults and children have argued that the most relevant temperamental factors might be: social engagement, such as compliance (Drake, 2010; Gudjonsson, 1983; Mastroberardino & Marucci, 2013; Shapiro & Purdy, 2005), shyness (Chae & Ceci, 2005; Endres, Poggenpohl, & Erben, 1999; Pozzulo, Coplan, & Wilson,

2005; Roebers & Schneider, 2001a; Shapiro, Blackford, & Chen, 2005; Young et al., 2003), approach-withdrawal (Burgwyn-Bailes, Baker-Ward, Gordon, & Ornstein, 2001; Gordon et al., 1993; Imhoff & Baker-Ward, 1999; Shapiro, 2006; Shapiro et al., 2005), adaptability (Geddie et al., 2000; Gordon et al., 1993; Shapiro, 2006; Shapiro et al., 2005) and emotionality (Burgwyn-Bailes et al., 2001; Chae & Ceci, 2005; Geddie et al., 2000; Gordon et al., 1993; Gudjonsson, 1983; Imhoff & Baker-Ward, 1999; McGroarty & Thomson, 2013; Raju et al., 1999; Shapiro, 2006). In fact these temperamental characteristics impact on an individual's adjustment to the interrogation context and consequently to the interviewee's memory accuracy and suggestibility (Shapiro & Purdy, 2005).

However, it is important to underline that the pattern of findings in relation to these variables is not consistent, since most of the studies failed to detect relationships with memory and suggestibility. Nevertheless, it is reasonable to assume that children's stable behavioural style has an impact on interrogative suggestibility, a phenomenon that involves socially mediated processes, in addition to purely cognitive ones. Consequently, it may be considered important to understand which temperamental factors are involved.

6.1. STUDY 8. The role of temperament on children's suggestibility

This study aimed to provide new data concerning the developmental changes in memory recall and interrogative suggestibility in school-aged children. Further, it aimed also to investigate whether individual-differences in temperamental traits are related to children's memory and interrogative suggestibility. The individual traits considered are measures of social engagement, inhibition to novelty, emotionality and self-regulation. The emotionality trait is considered in positive and negative valence separately.

It was hypothesised that temperamental traits interact in different ways for determine memory performance and vulnerability to suggestibility. In particular, it was

expected that social orientation, and positive emotionality and attention may facilitate the memory accuracy and fluency and they may be protective factors against interrogative suggestibility. On the contrary, it was hypothesised that inhibition to novelty, negative emotionality, activity, compliance and anxiety/shyness could have a negative effect on event memory and may reveal risk factors to succumb to interrogative suggestibility.

6.1.1. Method

Participants

A total of 366 children (179 males) ranging in age from 5.99 to 10.16 years (M = 8.09; SD =1.16) were recruited from eight primary schools in Italy[25]. The children were predominantly from Italian households (77.9%) and came from three strata of middle socio-economic backgrounds: lower-middle 28%, middle 45.2 % and upper-middle 26.8%. SES was derived from information provided by parents about their education and occupation.

All of the children included in the sample were intellectually within the normal range, as shown by their scores in the Raven's Coloured Progressive Matrices (CPM, Raven, 1998), all of which were above the 25th Percentile.

Instruments

Suggestibility

The Bicocca Suggestibility Scales-B Form, long version (BISS-B32), was used to investigate the children's memory and their vulnerability to interrogative suggestibility. The following measures were considered:

[25] Chiara Spinelli, Cecilia Giordano, Melissa Martinelli, Alessia Corticelli, Daniela Floria, Alessia Rota, Valeria Brandano, and myself collected the data.

Free Recall = the amount of information correctly reported during the immediate recall of the story (maximum = 49).

Yield = the number of leading questions to which the child has given the wrong answer during the administration of the first series of questions (maximum = 16).

Shift = the number of questions to which the child has changed his or her response after receiving the negative feedback (maximum = 32).

Temperament

The *Italian Questionnaires of Temperament* (QUIT; Axia, 2002) is an Italian battery of multidimensional questionnaires designed to assess child temperament, from a month to eleven years. These questionnaires assess six dimensions: Social Orientation, Inhibition to Novelty, Activity, Positive Emotionality, Negative Emotionality, and Attention. *Social Orientation* indicates the level of interest in, and attention to, social stimuli and to other people. High scores on this scale indicate an inclination to be very friendly with other people. *Inhibition to Novelty* is a measure of the child's reaction to novelty. High scores indicate a strong tendency to inhibit responses to novel stimuli. The latter four scales of the QUIT may be considered as measures of child's self-regulation: *Activity*, may be considered as a measure of the energy and motor regulation of the child. High scores indicate a strong tendency to being in movement and a lack of motor self-regulation. *Positive Emotionality*, where high scores indicate prevalent experiences of positive mood. *Negative Emotionality*, where high scores indicate prevalent experiences of negative mood. *Attention*, may be intended as a measure of cognitive regulation of the child and refers to individual differences in attention getting and attention holding. High scores indicate a good ability to concentrate and maintain attention while involved in a task.

Parents are required to rate each item from 1 ("not characteristic or typical of the child") to 6 ("very characteristic or typical of the child"); the mean score is considered for each scale.

The *Conners Parents Rating Scale-Revised* (CPRS-R; Conners, 2001) is a battery of multidimensional questionnaires which are useful in assessing the presence of a large spectrum of psychological problems in 3-17 year old subjects, regarding cognition, affectivity and social functioning by different subscales. However, in our study we considered only two subscales: Oppositional and Anxious-Shy. Two measures were derived from CPRS-R:L (Conners, 2001) filled-in by the parents, and were considered as indicators of child's social engagement: Compliance and Anxiety-Shyness. The first was derived from the Oppositional scale (10 items). Low scores on the Oppositional scale indicate compliance with social rules, while high scores indicate a propensity to violate them and to have trouble with authority. The second measure was derived from the Anxious-Shy scale (8 items). High scores on the Anxious-Shy scale reveal more preoccupations and fears than peers, higher sensitivity to criticism, anxiety in unfamiliar situations and reserve.

Procedure

Participants were individually tested in a quiet room at their school by one interviewer for one hour. The study was presented as a project on memory ability.

The parents filled in two questionnaires: the form for children aged 7-11 years was The Italian Questionnaires of Temperament (QUIT, Axia, 2002), and the long form of the Conners' Rating Scales Revised for Parents (CPRS-R:L, Conners, 2001).

6.1.2. Results

Preliminary analyses showed that memory and suggestibility measures did not differ in relation to gender, so the subsequent analyses were conducted on data regardless of gender.

The descriptive statistics of the memory, suggestibility and psycho-social measures are shown in Table 6.1. An absence of questionnaires filled in by parents reduced the number of participants to 356 for the data involving the Italian Questionnaires of Temperament and to 125 for the Conners' Rating Scales Revised for Parents.

Table 6.1. Descriptive statistics of memory, suggestibility and temperamental traits.

	Range			M	SD	Skewness	Kurtosis
Free recall	0.00	-	41.00	21.92	7.61	-0.37	-0.11
Yield	0.00	-	15.00	5.99	3.81	0.09	-1.11
Shift	0.00	-	26.00	5.80	4.48	1.78	4.46
Social Orientation	0.00	-	6.00	4.72	0.71	-1.27	5.23
Inhibition to Novelty	1.00	-	6.00	2.59	0.90	0.67	0.28
Activity	1.74	-	5.80	3.86	0.90	-0.09	-0.75
Positive Emotionality	2.13	-	6.25	4.64	0.74	-0.88	0.71
Negative Emotionality	1.00	-	4.63	2.49	0.73	0.49	0.13
Attention	1.77	-	5.89	4.15	0.76	-0.38	0.16
Compliance	37.00	-	93.00	52.00	11.09	1.35	2.46
Anxiety/ Shyness	36.00	-	83.00	52.33	10.54	0.72	-0.15

In order to check the relationship between age, memory and suggestibility, Pearson's correlations were conducted. In terms of free recall, the analyses showed a positive relation with age (free recall: $r = .65$, $p < .001$). Regarding suggestibility, the analyses showed that, overall, younger children proved to be more suggestible than older children (Yield: $r = -.36$, $p < .001$; Shift: $r = -.41$, $p < .001$).

To investigate the relationships between the memory and suggestibility scores and psycho-social measures, partial correlations were used controlling for age. Table 6.2 displays the partial correlations for the variables examined in this study.

Table 6.2. Partial correlations of memory and suggestibility with temperamental traits.

	1.	2.	3.
1. Free recall			
2. Yield	-.09		
3. Shift	-.04	.30***	
4. Social Orientation	.23***	-.14**	.00
5. Inhibition to Novelty	-.25***	.08	.13*
6. Activity	-.01	-.15**	-.14*
7. Positive Emotionality	.17**	-.21***	-.11*
8. Negative Emotionality	-.03	.10	.11*
9. Attention	.17**	-.07	-.10
10. Compliance	.05	.02	.07
11. Anxiety/Shyness	-.16	.07	.14

p=.05; *p<.05; **p<.01; ***p<.001.

In terms of memory, Free Recall was positively correlated to Social Orientation, Positive Emotionality and Attention, while it was negatively correlated to Inhibition to Novelty. In terms of suggestibility, Yield and Shift scores were strongly related. Yield was negatively correlated with Social Orientation, Activity and Positive Emotionality. Shift was positively correlated with Inhibition to Novelty and to Negative Emotionality, while it was negative correlated with Activity and Positive Emotionality. There were no other significant associations (all *p*s >.05).

To assess whether the temperament factors that were correlated with our outcome measures contributed independent predictive power, three hierarchical regression analyses were conducted. First, children's free recall score was considered as dependent measure (see Table 6.3). Age was entered at Step 1 and accounted for 42% of the variation in free recall score. This model was statistically significant, F (1, 353) = 253.37, p < .001. Then, Social Orientation, Positive Emotionality, Attention, and Inhibition to Novelty were simultaneously added at Step 2. By adding temperamental traits, the model accounted for additional 5% of the variation in free recall score (final R^2 = .47), and it was statistically significant F (5, 349) = 61.07, p < .001. Social Orientation (sr^2 = .01)

and Inhibition to Novelty (sr^2 = .01) emerged as significant predictors of free recall score.

Table 6.3. Hierarchical regression analyses predicting free recall score from age and temperamental traits.

		B	*SE B*	β
Step 1				
	Constant	-12.46	2.18	
	Age	4.25	0.27	.65***
Step 2				
	Constant	-16.64	4.16	
	Age	4.17	0.26	.63***
	Social Orientation	1.16	0.49	.11*
	Inhibition to Novelty	-1.06	0.41	-.13*
	Positive Emotionality	0.17	0.48	.02
	Attention	0.32	0.47	.03

Note: R^2 = .42 for Step 1, ΔR^2 = .05 for Step 2 ($p < .001$). Final R^2 = .47.
*$p < .05$, **$p < .01$, ***$p < .001$

To investigate whether temperament factors contributed independent predictive power to a model comprising age, a hierarchical regression analysis was then conducted with Yield score as the dependent variable (see Table 6.4). Age was entered at Step 1, and accounted for 13% of the variation in yield score. This model was statistically significant, F (1, 354) = 52.29, $p < .001$. Temperamental traits were added at Step 2: Social Orientation, Activity and Positive Emotionality contributed to a statistically significant change, F (4, 351) = 19.16, $p < .001$ and the model accounted for additional 5% of the variation in yield score (final R^2 = .18). Both Activity (sr^2 = .01) and Positive Emotionality (sr^2 = .02) emerged as independent predictors of children's tendency to Yield to misleading questions. Social Orientation did not contribute significantly to the equation.

Finally, to investigate whether temperament factors contributed independent predictive power to a model comprising age and Yield score, a hierarchical regression analysis was then conducted with Shift score as the dependent variable (see Table 6.5). Age was entered at Step 1 and accounted for 17% of the variation in shift score. This model was statistically significant, F (1, 354) = 72.31, $p < .001$. Yield was added at Step 2 and accounted for additional 7% of the variation in shift score, and it was statistically significant F (2, 353) = 56.54, $p < .001$. Then, Inhibition to Novelty, Activity, Positive Emotionality and Negative Emotionality were added at Step 3. By adding temperamental traits, the model accounted for additional 1% of the variation in shift score (final R^2 = .25), and it was statistically significant F (3, 352) = 39.36, $p < .001$. However, only Inhibition to Novelty (sr^2 = .01) emerged as independent predictors of children's tendency to shift their first answer after negative feedback. Activity, Positive and Negative Emotionality did not contribute significantly to the equation and were removed from the model.

Table 6.4. Hierarchical regression analyses predicting Yield score from age and temperamental traits.

		B	*SE B*	β
Step 1				
	Constant	15.53	1.33	
	Age	-1.18	0.16	-0.36***
Step 2				
	Constant	22.67	2.05	-0.38***
	Age	-1.26	0.16	-0.04
	Social Orientation	-0.20	0.30	-0.11
	Activity	-0.47	0.21	-0.16*
	Positive Emotionality	-0.81	0.28	-0.38**

Note: R^2 = .13 for Step 1, ΔR^2 = .05 for Step 2 ($p < .001$). Final R^2 = .18.

*$p < .05$, **$p < .01$, ***$p < .001$

Table 6.5. Hierarchical regression analyses predicting Shift score from age, Yield and temperamental traits.

		B	*SE B*	β
Step 1				
	Constant	18.70	1.53	
	Age	-1.60	0.19	-.41***
Step 2				
	Constant	13.41	1.73	
	Age	-1.19	0.19	-.31***
	Yield	0.34	0.06	.29***
Step 3	Constant	12.26	1.81	
	Age	-1.19	0.19	-.31***
	Yield	0.33	0.06	.28***
	Inhibition to Novelty	0.46	0.23	.09*

Note: R^2 = .17 for Step 1, ΔR^2 = .07 for Step 2 (p < .001), ΔR^2 = .01 for Step 3 (p = .046). Final R^2 = .25.

*p < .05, **p < .01, ***p < .001

6.1.3. Discussion

This study confirmed the presence of pronounced age differences both in terms of recall ability and suggestibility. Older children performed better than their younger counterparts in free recall (Caprin et al., 2016; Chae & Ceci, 2005; Chae, Goodman, Eisen, & Qin, 2011b; Roebers & Schneider, 2001a), whereas younger children proved to be more suggestible, in line with previous research conducted with similar procedures (Caprin et al., 2016; Danielsdottir et al., 1993; Roma et al., 2011). This means that the changes occurring in the typical development with increasing age affect memory and suggestibility.

The findings of the present study showed that not all the temperamental factors which were investigated, were related to memory recall and suggestibility. Regarding memory, the results partially confirmed the initial hypothesis. Compliance did not

correlate with the free recall tasks, while Social Orientation was positively related to Free recall and it also contributed to predict the children's performance on the free recall task. These findings are only partially in line with previous studies. In fact, Compliance does not seem to be related to memory as suggested by the literature (Burgwyn-Bailes et al., 2001; Muir-Broaddus, King, Downey, & Petersen, 1998), and in terms of Anxiety-Shyness, contrary to most studies, which reported a negative association with recall tasks (Chae & Ceci, 2005; Gordon et al., 1993; Muir-Broaddus et al., 1998; Roebers & Schneider, 2001a), no associations with free recall were found in the present study. This result would support the hypothesis that shy children are less accurate than non-shy ones. However, many studies did not detect this association (Burgwyn-Bailes et al., 2001; Geddie et al., 2000; Imhoff & Baker-Ward, 1999; Liebman et al., 2002; Pozzulo, Crescini, Lemieux, & Tawfik, 2007; Rocha, 2003). The positive relation found between Social Orientation and free recall would suggest that more sociable children tend to report more information than more reserved children, because their social skills make them feel at ease in a social situation such as a conversation (Roebers & Schneider, 2001a). Few previous studies that investigated Social Orientation, did not found relation with memory (Burgwyn-Bailes et al., 2001; Liebman et al., 2002; Ward & Loftus, 1985), except Rocha (2003), who found a positive association with free recall, the same as us.

Inhibition to Novelty was negatively related with free recall, so children with higher levels of inhibition are less accurate in their memory performances, as suggested by Schneider and Sodian (1991). This finding is in line with the work of Liebman et al. (2002) on adults, but most of the studies failed to detect this relationship (Burgwyn-Bailes et al., 2001; Imhoff & Baker-Ward, 1999), except Geddie et al. (2000) who found a positive association with memory accuracy.

Furthermore, most of the previous studies that considered the Emotionality trait, did not detect associations with memory performance, despite expectations to the contrary (Burgwyn-Bailes et al., 2001; Chae & Ceci, 2005; Geddie et al., 2000; Imhoff & Baker-

Ward, 1999; Liebman et al., 2002). Conversely, in Study 8 a positive association between Free Recall and Attention was found. This result is in line with a previous study with preschoolers (Benedan, Zajac, McFarlane, & Powell, 2018) even if, surprisingly, in Study 8 no significant relationship between interrogative suggestibility scores and Attention was found.

Regarding interrogative suggestibility measures, results confirmed some of the expectations. The Social-Engagement measures were not related to interrogative suggestibility. Despite the fact that it is often speculated that the obedience and compliance of children might be a factor involved in individual differences to suggestibility (Eisen, Gomes, Lorber, Perez, & Uchishiba, 2013; Schacter, Kagan, & Leichtman, 1995), only a few studies have investigated this issue and most of them failed to detect any relationships (Muir-Broaddus et al., 1998; Shapiro & Purdy, 2005). However, as shown in Table 6.1, the mean score was around the 50th percentile and the data dispersion was very low, in fact 80% of the children's measures were between the 40th and the 60th percentile, hence it could be that our measures reduced the chances of finding a significant relationship, so this issue warrants further investigation.

The literature concerning the relationship between anxiety-shyness and suggestibility in children report inconsistent findings. Despite the reasonable hypothesis that anxiety-shyness might increase the susceptibility to suggestion (Chae & Ceci, 2005; Roebers & Schneider, 2001a), few studies found significant relationships (Endres et al., 1999), whereas most studies found no associations (Burgwyn-Bailes et al., 2001; Chae & Ceci, 2005; Muir-Broaddus et al., 1998; Rocha, 2003; Roebers & Schneider, 2001a; Shapiro & Purdy, 2005; Young et al., 2003). However, also in terms of this measure it is likely that the low variability in this measure reduced the chances of finding a significant relationship, because 70% of the children's measures were between the 40th and the 60th percentile, so this issue warrants further investigation.

The temperamental dimension of Social Orientation was negatively related to the Yield score. This would mean that a high level of interest to other people would be associated with a reduced vulnerability to misleading questions. Social Orientation is conceptually related to Shyness, but also regarding this issue the findings of previous studies are inconsistent. Some studies did not find relationships with suggestibility (Geddie et al., 2000; Imhoff & Baker-Ward, 1999), others found a positive relationship (Burgwyn-Bailes et al., 2001) and still others found a negative relationship (Rocha, 2003).

Regarding Inhibition to Novelty, we found that it was positively related to Shift but not to Yield. This finding suggests that the children who are more inhibited may be more susceptible to negative feedback and change their responses more than children with low levels of Inhibition to Novelty. Also Kagan, Snidman, Arcus, and Reznick (1994) found that inhibited children were reluctant to oppose adult requests and they speculated that it may be relevant in interrogative contexts, increasing children's susceptibility to suggestibility. However, most previous research which investigated a similar dimension of adaptability failed to detect a relationship to Yield in children, but contrary to the results of Study 8, these studies did not consider Shift (Burgwyn-Bailes et al., 2001; Geddie et al., 2000; Imhoff & Baker-Ward, 1999; Shapiro & Purdy, 2005). Shapiro et al. (2005) studied the vulnerability to misleading questions in distinguishing different categories of details concerning an event represented by a video-story, in children of 6-7 years, and 9-10 years, and in adults. They found significant associations between adaptability and suggestibility for only a few of the categories, and in particular they detected only one positive correlation among the five calculations in 9/10 years old children, and two among the five calculations in 6/7 years old children, but the signs were contrasting, i.e. one positive and one negative.

Activity was negatively associated with both Yield and Shift. This result was unexpected. It would mean that an increased motor activity is associated to a decreased vulnerability to suggestion and a decreased tendency to change answer after negative

feedback. Moreover, this temperamental trait was a significant predictor of the Yield score, over and above age.

Interestingly a negative correlation between Positive Emotionality and Yield was found. Distinguishing between the opposite valence of emotionality, Positive Emotionality correlated negatively with Yield and Shift, while Negative Emotionality was positively associated with Shift only. Moreover, the relative contribute of Positive Emotionality on Yield score was significant over and above the age effect. Previous studies with children have generally investigated Emotionality as a global dimension, without distinguishing positive and negative emotions and the findings are controversial. Chae & Ceci (2005) found a positive correlation between Emotionality and yielding to misleading questions in school-age children. Also Shapiro et al. (2005) found that Emotional Intensity was related to suggestibility in children, but their findings were contrasting: they found significant correlations only for 2 of the 5 calculations both in 6/7 and 9/10 years old children, but in both the groups the signs of the correlations were contrary, that is one positive and the other one negative. However, most of previous studies failed to detect relationships with suggestibility (Burgwyn-Bailes et al., 2001; Geddie et al., 2000; Imhoff & Baker-Ward, 1999).

In conclusion, it is widely documented that the accuracy of children's reports hinges largely on the way in which they are interviewed, but individual factors mediate the impact of the interview, especially when suggestive (Bruck & Melnyk, 2004). The investigations on the role of temperamental factors are limited, compared with those on cognitive factors, but certain characteristics, such as social orientation, inhibition to novelty, activity, emotionality and attention, impact the adjustment of children to the interview context and consequently their memory accuracy and suggestibility (Shapiro & Purdy, 2005). Further research is needed to identify the most important temperamental factors which contribute to a vulnerability to interrogative suggestibility.

CHAPTER 7

General discussion and conclusions

In conclusion, the studies reported in this thesis sought to examine the validity and usefulness of the BISS within the context of children's eyewitness memory. The first aim of this thesis was to examine the reliability and validity of the BISS within some Italian samples of school-aged children. The BISS was recently developed by Caprin and myself at University of Milano-Bicocca, specifically for use with 6- to 11-year-old Italian children. As outlined by Hambleton and Patsula (1998) the development of a new test may rise validity over the validity level of the adapted test. This could be the case of the Bicocca Suggestibility Scales, since suggestibility is a construct that largely rely on language.

This test battery has been created with the purpose of having a reliable instrument to discriminate between children who have different levels of suggestibility. Because it relies on a memory paradigm, it has two parallel versions (BISS-A and BISS-B), to administer the test more than once to the same child.

Study 1 to Study 4 of the present thesis examined the psychometric properties of the BISS. First, the internal consistency and test-retest reliability of the memory and suggestibility measures were evaluated.

Specifically, it was examined the degree to which children's performance in a given context was stable over time. Children were administered the parallel versions of the BISS on two separate occasions: after two weeks in Study 1 and after two years in Study 2. The results of these studies would provide some support to the hypothesis that suggestibility is a stable trait-like tendency.

Study 3 examined the construct validity of the BISS, with special reference to the concurrent validity. This was done by comparing children's performance on the BISS-

A32 with their performance on the Bonn Test of Statement Suggestibility (Endres, 1997; Rossi et al., 2011), which is a standardised test of suggestibility developed for school-aged children.

In Study 4, a factor analysis was conducted to assess the structure of the BISS.

Results of the first four studies allow to conclude with enough confidence that the Bicocca Suggestibility Scales is a valid and reliable tool to assess individual differences in interrogative suggestibility in Italian school-aged children.

At least one dimension of suggestibility, i.e. children's willingness to yield to misleading questions, satisfied all the psychometric criteria. Indeed, the Yield score, which represents the child's tendency to accept the suggestion by providing a wrong answer to misleading questions, resulted to have a good internal consistency and good test-retest reliability. Some concerns may be addressed to the Shift score. Major psychometric and conceptual problems have been associated with measuring the notion of Shift with the GSS, and therefore, Gignac and Powell (2009) recommended to limit the use of the GSS to the Yield subscale. The BISS protocol structure has been created to overtake this limit of the GSS, to have a more reliable measure of the Shift score. However, the results of the present thesis showed that some methodological issues are still present. Indeed, most of the children did not change their first answer after receiving negative feedback. Therefore, the high presence of constants in the items used to compute the Shift score determines a scarce variability. However, internal consistency of the Shift score has been found to be satisfactory. So, even if caution should be used, the Shift score may still be applied in future research with the BISS.

The development of a valid standardised test to assess interrogative suggestibility in children was oriented to both the research and the applied context. From the research point of view, a similar tool could enhance studies aimed to deepen understanding of the suggestibility phenomenon.

From an applied point of view, this instrument could be used as a diagnostic tool to evaluate the reliability and credibility of an account of abuse made by a child. This could help investigative interviewers and child protection workers, in determining the child's susceptibility to suggestion. However, it is important to state that this does not mean to use this test to exclude a child from the legal process. The aim would be to help practitioners to consider the particular needs and abilities of each child when interviewing children (McFarlane & Powell, 2002).

Assessing a child's vulnerability to misleading questions and interpersonal pressure may be considered as an imperfect solution to the problem of the reliability of child's testimonies (Endres, 1997). Indeed, it would be better to avoid at all suggestive influences during the investigative process, being them leading or misleading questions rather than question repetition or interviewer's feedback. Most recent child investigative interviewing protocols already warn interviewers against the use of misleading questions (Brown et al., 2013; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007). Open-ended questions should be the first option when questioning children (Bull, 2010), because this type of question allows to obtain longer responses even with younger children (Sternberg et al., 1996) and maximise child credibility (Feltis, Powell, Snow, & Hughes-Scholes, 2010). Furthermore, interviewer's adherence to open-ended questions would reduce the negative consequences of interviewer confirmation bias (Powell, Hughes-Scholes, & Sharman, 2012). However, the use of open-ended questions is not a panacea and some issues have been raised due to the presence of inconsistencies in children's reports (Pichler & Sharman, 2017). In addition, despite the best-practice guidelines, research has shown that adults interviewing children tend to use only a small percentage of open-ended question and a high proportion of specific closed-questions. This has been proved for police interviewers and social workers (Powell, Fisher, & Wright, 2005; Sternberg et al., 2001), but also for lawyers and judges in countries where

children are subjected to court questioning (Powell, Westera, Goodman-Delahunty, & Pichler, 2016; Stolzenberg & Lyon, 2014).

Therefore, as far as it is not possible to shield children from the types of questions that could contaminate their reports, the next best option is to identify the children who are likely experience particular difficulty answering them accurately (Benedan, Zajac, et al., 2018). With the aim of understanding which individual difference factors could be associated to a heightened suggestibility, the last four studies of the present thesis were conducted. These studies had the dual purpose of examining the discriminant validity of the BISS, by proving that the BISS can satisfactorily discriminate interrogative suggestibility from cognitive and temperamental factors.

Many cognitive and social variables have been linked with interrogative suggestibility without arriving to a complete picture of the phenomenon (for a review, see Bruck & Melnyk, 2004). Age has been considered the main one, and marked age differences in suggestibility are apparent within the empirical literature. In general, adults are less suggestible than children, and older children are less suggestible than younger children (Ceci & Bruck, 1993), being preschoolers the most vulnerable to the effects of suggestion (Bjorklund et al., 2000; Peterson & Biggs, 1997). In the present thesis, typical age effects were found in interrogative suggestibility. Suggestibility has been studied in samples of school-aged children, by showing that this age is still critical for the vulnerability to suggestion and that developmental changes determining typical abilities of this developmental phase exert a great impact on the vulnerability to suggestibility.

Furthermore, this thesis provided new data about the complex relationship between memory and suggestibility for the same event. Indeed, all the suggestibility measures were related to free recall (with the exception of the post-event misinformation effect), providing evidence in support of the trace-theory (Ceci & Bruck, 1993).

Moreover, new data was provided on the post-event misinformation effect, measured by free recall procedure. The various forms of suggestibility were associated

with each other. This finding suggests that the hypothesis that children who are more prone to suggestion under some circumstances are also in others, supporting the view that even if suggestibility is a multidimensional concept, it may be reasonably thought of as a trait. The BISS introduced a new procedure to evaluate the interference of misleading information on children's memory. It deserves to be examined further in future research.

Study 5 did not reveal the existence of a relationship between suggestibility and cognitive functioning, except for working memory, in line with most previous studies (Bruck & Melnyk, 2004). This relationship has been deepened in Study 6, where the relative contribution of executive functioning and theory of mind was investigated.

The results of the Studies 6 and 7 provide evidence to support the hypothesis that theory of mind is a protective factor against vulnerability to interrogative suggestibility. Indeed, children with higher scores in the Vocal Test of Mental States performed better when answering misleading questions of the BISS and of the interviewing protocol about the video. Moreover, the susceptibility to misleading questions was affected by the interviewer's style. Indeed, when the interviewer's attitude was abrupt, children tended to be more vulnerable to suggestion than children interviewed by a friendly adult. This result confirms the findings of previous studies about the role of interviewer's support on the quality of children's reports (Carter et al., 1996; Davis & Bottoms, 2002; Goodman et al., 1991; Quas et al., 2005). However, contrary to our expectations, no significant interactions between children's theory of mind and interviewer demeanour were observed.

Finally, Study 8 investigated the role of temperamental factors in predicting children's performance in memory and suggestibility tasks. Free Recall was positively correlated to Social Orientation, Positive Emotionality and Attention, while it was negatively correlated to Inhibition to Novelty. However, only Social Orientation and Inhibition to Novelty contribute significantly in predicting the free recall score after controlling for age. Yield was negatively correlated with Social Orientation, Activity and

Positive Emotionality, but only these latter measures were significant predictors of the answers to misleading questions. Shift was positively correlated with Inhibition to Novelty and to Negative Emotionality, while it was negative correlated with Activity and Positive Emotionality but these associations had low effect sizes and did not add explicative value to the model comprising age and the yield factor.

Overall, a complex framework emerged, by indicating the several cognitive and temperamental factors impact the adjustment of children to the interview context. Further research is needed to better understand the simultaneous role of all these factors in contributing to a vulnerability to interrogative suggestibility.

Being the BISS a newly developed scale, further research is clearly needed to complete the validation process of the instrument and to explore the generalisability of the findings of the current thesis about the use of the BISS in different contexts. Moreover, future studies should examine the issue of whether the BISS is sensitive enough to discriminate children coming from special population, for instance children with mild or moderate intellectual disabilities (Gudjonsson & Henry, 2003; Henry & Gudjonsson, 2007; Young et al., 2003), or children exposed to negative experiences such as abuse and neglect (Benedan, Powell, et al., 2018; Curci et al., 2016; Vagni et al., 2015).

Predictive validity of the BISS should also be assessed. Some preliminary data about the predictive validity of the BISS have been provided in Study 7. Indeed, the BISS Yield score contributed significantly in predicting the answers to misleading questions about the video, over and above the age effect. The regression model showed that the BISS-A32 Yield measure was the most important predictor of the accuracy of children's responses in the interview about the video, over and above age. This result would suggest that BISS Yield scores are good predictors of children's performance on a similar task, even more than chronological age. Although this finding opens the possibility of introducing the Bicocca Suggestibility Scales as a forensic tool in the assessment of child witnesses, it is premature to determine the actual utility and predictive validity of this

instrument. The results of the current study should be replicated in future research by introducing a comparison with suggestibility for an experienced event.

Furthermore, future studies should address some topics that have not been dealt with in depth. For instance, the post-event misinformation effect detected with the BISS procedure is quite unusual in the suggestibility research. Indeed, it is assessed by analysing the spontaneous recall of suggested information rather than use a recognition task as in the standard misinformation procedure (see Loftus, 2005).

Finally, as shown in Study 5, suggestibility measures were differentially related to spontaneous confabulations. The concept of confabulations meant as spontaneous false memories should be deepened and the association with the different types of suggestibility warrants further exploration.

REFERENCES

Agnew, S. E., & Powell, M. B. (2004). The effect of intellectual disability on children's recall of an event across different question types. *Law and Human Behavior*, *28*(3), 273–294.

Alexander, K. W., Goodman, G. S., Schaaf, J. M., Edelstein, R. S., Raju, S., & Shaver, P. R. (2002). The role of attachment and cognitive inhibition in children's memory and suggestibility for a stressful event. *Journal of Experimental Child Psychology*, *83*(4), 262–290. https://doi.org/10.1016/S0022-0965(02)00149-2

Andrade, J. (2001). Working memory in context. *Psychology, Hove, Sussex*.

Asendorpf, J. B. (1989). Individual, differential, and aggregate stability of social competence. In B. H. Schneider, G. Attili, J. Nadel, & R. P. Weissberg (Eds.), *Social Competence in Developmental Perspective* (pp. 71–86). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-2442-0_4

Asendorpf, J. B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality*, *6*, 103–117.

Aveling, F., & Hargreaves, H. L. (1921). Suggestibility with and without prestige in children. *British Journal of Psychology*, *11*, 53–75.

Axia, G. (2002). *QUIT. Questionari italiani del temperamento*. Edizioni Erickson.

Bain, S. A., & Baxter, J. S. (2000). Interrogative suggestibility: The role of interviewer behaviour. *Legal and Criminological Psychology*, *5*, 123–133. https://doi.org/10.1348/135532500168029

Bain, S. A., Baxter, J. S., & Fellowes, V. (2004). Interacting influences on interrogative suggestibility. *Legal and Criminological Psychology*, *9*(2), 239–252. https://doi.org/10.1348/1355325041719419

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford university press.

Baxter, J. S. (1990). The suggestibility of child witness: a review. *Applied Cognitive Psychology*, *4*, 393.

Baxter, J. S., & Boon, J. C. W. (2000). Interrogative suggestibility: the importance of being earnest. *Personality and Individual Differences*, *28*(4), 753–762. https://doi.org/10.1016/S0191-8869(99)00136-1

Baxter, J. S., Jackson, M., & Bain, S. A. (2003). Interrogative suggestibility: Interactions between interviewees' self-esteem and interviewer style. *Personality and Individual Differences*,

35(6), 1285–1292. https://doi.org/10.1016/S0191-8869(02)00349-5

Bayley, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. *The Journal of Genetic Psychology*, *75*, 165–196.

Benedan, L., Powell, M. B., Zajac, R., Lum, J. A. G., & Snow, P. C. (2018). Suggestibility in neglected children: Relationship with intelligence, language and social skills. *Child Abuse & Neglect, 79,* 51-60. https://doi.org/10.1016/j.chiabu.2018.01.005

Benedan, L., Zajac, R., McFarlane, F., & Powell, M. B. (2018). Temperamental risk factors for interrogative suggestibility in preschoolers. *Manuscript Submitted for Publication*.

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior*, *19*(1), 79–88. https://doi.org/10.1007/BF01499074

Bernstein, D. M., & Loftus, E. F. (2009). How to tell if a particular memory is true or false. *Perspectives on Psychological Science*, *4*(4), 370–374. https://doi.org/10.1111/j.1745-6924.2009.01140.x

Bianco, A., & Curci, A. (2015). Measuring interrogative suggestibility with the Italian version of the Gudjonsson Suggestibility Scales (GSS): Factor structure and discriminant validity. *Personality and Individual Differences*, *82*, 258–265. https://doi.org/10.1016/j.paid.2015.03.035

Binet, A. (1900). *La suggestibilité*. Paris: Schleicher Frères.

Binet, A. (1904). La science du témoignage. *L'Année Psychologique*, *11*, 128–136.

Binet, A., & Henri, V. (1894). De la suggestibilité naturelle chez les enfants. *Revue Philosophique de La France et de l'Étranger*, *38*, 337–347.

Bjorklund, D. F., Cassell, W. S., Bjorklund, B. R., Brown, R. D., Park, C. L., Ernst, K., & Owen, F. A. (2000). Social demand characteristics in children's and adults' eyewitness memory and suggestibility: The effect of different interviewers on free recall and recognition. *Applied Cognitive Psychology*, *14*(5), 421–433. https://doi.org/http://dx.doi.org/10.1002/1099-0720(200009)14:5<421::AID-ACP659>3.0.CO;2-4

Bland, J. M., & Altman, D. G. (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society*, *32*(3), 307–317.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160.

https://doi.org/10.1191/096228099673819272

Bland, J. M., & Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*, *47*(8), 931–936. https://doi.org/10.1016/j.ijnurstu.2009.10.001

Bottoms, B. L., Quas, J. A., & Davis, S. L. (2007). The influence of the interviewer-provided social support on children's suggestibility, memory, and disclosures. In M.-E. Pipe, M. E. Lamb, Y. Orbach, & A. C. Cederborg (Eds.), *Child sexual abuse: Disclosure, delay and denial* (pp. 135–157). Erlbaum Hillsdale, NJ.

Brady, M. S., Poole, D. A., Warren, A. R., & Jones, H. R. (1999). Young children's responses to yes-no questions: patterns and problems. *Applied Developmental Science*, *3*(1), 47–57. https://doi.org/10.1207/s1532480xads0301_6

Brainerd, C. J., & Howe, M. L. (1982). Stages-of-learning analysis of developmental interactions in memory, with illustrations from developmental interactions in picture-word effects. *Developmental Review*, *2*(3), 251–273. https://doi.org/10.1016/0273-2297(82)90014-4

Brainerd, C. J., & Mojardin, A. (1998). Children's and adults' spontaneous false memories: Long-term persistence and mere-testing effects. *Child Development*, *69*(5), 1361–1377.

Braun, K. A., Loftus, E. F., & Ellis, R. (2002). Make my memory: how advertising can change our memories of the past. *Psychology & Marketing*, *19*(1), 1–23. https://doi.org/10.1017/CBO9781107415324.004

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, *18*(6), 906–921. https://doi.org/10.1037/0012-1649.18.6.906

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2008). Theory-of-mind development influences suggestibility and source monitoring. *Developmental Psychology*, *44*(4), 1055–1068. https://doi.org/10.1037/0012-1649.44.4.1055

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: a dimensional and developmental study. *Developmental Neuropsychology*, *26*(2), 571–593. https://doi.org/10.1207/s15326942dn2602_3

Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M.-E., Orbach, Y., & Wolfman, M. (2013). The NICHD Investigative Interview Protocol: An analogue study, *19*(4), 367–382. https://doi.org/10.1037/a0035143

Bruck, M., Ceci, S. J., & Melnyk, L. (1997). External and internal sources of variation in the

creation of false reports in children. *Learning and Individual Differences*, *9*(4), 289–316. https://doi.org/10.1016/S1041-6080(97)90011-X

Bruck, M., & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology*, *18*(8), 947–996. https://doi.org/10.1002/acp.1070

Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology*, *15*, 5–23. https://doi.org/10.1348/014466509X440160

Burgwyn-Bailes, E., Baker-Ward, L., Gordon, B. N., & Ornstein, P. A. (2001). Children's memory for emergency medical treatment after one year: The impact of individual difference variables on recall and suggestibility. *Applied Cognitive Psychology*, *15*, S25–S48. https://doi.org/10.1002/acp.833

Caprin, C., & Benedan, L. (2012). Presentazione del Bicocca Suggestibility Scales (BISS): un test per misurare la suggestionabilità interrogatoria in età scolare. In *Poster session presented at the Congresso AIAMC, XVI ed.* Pescara, Italy.

Caprin, C., Benedan, L., Ciaccia, D., Mazza, E., Messineo, S., & Piuri, E. (2016). True and false memories in middle childhood: the relationship with cognitive functioning. *Psychology Crime & Law*, *22*(5), 473–494. https://doi.org/10.1080/1068316X.2016.1168420

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*(6), 1105–1122. https://doi.org/10.1037/0012-1649.40.6.1105

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319. https://doi.org/10.1016/j.jecp.2004.01.002

Carter, C. A., Bottoms, B. L., & Levine, M. (1996). Linguistic and socioemotional influences on the accuracy of children's reports. *Law and Human Behavior*, *20*(3), 335–358.

Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, *113*(3), 403–439. https://doi.org/10.1037/0033-2909.113.3.403

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. American Psychological Association.

Chae, Y., & Ceci, S. J. (2005). Individual differences in children's recall and suggestibility: The effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology*, *19*(4), 383–407. https://doi.org/10.1002/acp.1094

Chae, Y., Goodman, G. S., Eisen, M. L., & Qin, J. (2011a). Event memory and suggestibility in abused and neglected children: Trauma-related psychopathology and cognitive functioning. *Journal of Experimental Child Psychology*, *110*(4), 520–538. https://doi.org/10.1016/j.jecp.2011.05.006

Cicchetti, D., Rogosch, F. A., Maughan, A., Toth, S. L., & Bruce, J. (2003). False belief understanding in maltreated children. *Development and Psychopathology*, *15*, 1067–1091. https://doi.org/10.1017.S0954579403000440

Clare, I. C. H., & Gudjonsson, G. H. (1993). Interrogative suggestibility, confabulation, and acquiescence in people with mild learning disabilities (mental handicap): Implications for reliability during police interrogations. *British Journal of Clinical Psychology*, *32*(3), 295–301. https://doi.org/10.1111/j.2044-8260.1993.tb01059.x

Clare, I. C. H., Gudjonsson, G. H., Rutter, S. C., & Cross, P. (1994). The inter-rater reliability of the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, *33*(3), 357–365. https://doi.org/10.1016/0191-8869(93)90197-B

Clarke-Stewart, K. A., Malloy, L. C., & Allhusen, V. D. (2004). Verbal ability, self-control, and close relationships with parents protect children against misleading suggestions. *Applied Cognitive Psychology*, *18*(8), 1037–1058. https://doi.org/10.1002/acp.1076

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Conners, C. K. (2001). *Conners' rating scales revised*. Multi-Health Systems, Incorporated.

Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (2007). *Variation in working memory*. Oxford University Press.

Curci, A., & Bianco, A. (2014). Gudjonsson Suggestibility Scales. Manuale d'uso. *Giunti OS Organizzazioni Speciali, Firenze*.

Curci, A., Bianco, A., & Gudjonsson, G. H. (2016). Verbal ability, depression, and anxiety as correlates of interrogative suggestibility in children exposed to life adversities. *Psychology, Crime & Law*, *23*(5), 445–458. https://doi.org/10.1080/1068316X.2016.1269901

Danielsdottir, G., Sigurgeirsdottir, S., Einarsdottir, H. R., & Haraldsson, E. (1993). Interrogative suggestibility in children and its relationship with memory and vocabulary. *Personality and Individual Differences*, *14*(3), 499–502. https://doi.org/10.1016/0191-8869(93)90323-U

Davis, S. L., & Bottoms, B. L. (2002). Effects of social support on children's eyewitness reports: A test of the underlying mechanism. *Law and Human Behavior*, *26*(2), 185–215.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17.

Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Psychology Press.

Doise, W., & Mugny, G. (1981). *Le développement social de l'intelligence* (Vol. 1). InterEditions Paris.

Drake, K. E. (2010). Interrogative suggestibility: Life adversity, neuroticism, and compliance. *Personality and Individual Differences*, *48*(4), 493–498. https://doi.org/10.1016/j.paid.2009.11.030

Eisen, M. L., Gomes, D. M., Lorber, W. G., Perez, C. I., & Uchishiba, H. (2013). Using an individual differences approach to examine two distinct types of suggestibility effects. *Applied Cognitive Psychology*, *27*(1), 2–11. https://doi.org/10.1002/acp.2864

Eisen, M. L., Qin, J., Goodman, G. S., & Davis, S. L. (2002). Memory and suggestibility in maltreated children: Age, stress arousal, dissociation, and psychopathology. *Journal of Experimental Child Psychology*, *83*(3), 167–212. https://doi.org/10.1016/S0022-0965(02)00126-1

Endres, J. (1997). The suggestibility of the child witness: The role of individual differences and their assessment. *Journal of Credibility Assessment and Witness Psychology*, *1*(2), 44–67.

Endres, J., Poggenpohl, C., & Erben, C. (1999). Repetitions, warnings and video: Cognitive and motivational components in preschool children's suggestibility. *Legal and Criminological Psychology*, *4*, 129–146. https://doi.org/http://dx.doi.org/10.1348/135532599167725

Feltis, B. B., Powell, M. B., Snow, P. C., & Hughes-Scholes, C. H. (2010). An examination of the association between interviewer question type and story-grammar detail in child witness interviews about abuse. *Child Abuse and Neglect*, *34*(6), 407–413. https://doi.org/10.1016/j.chiabu.2009.09.019

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.

Finnilä, K., Mahlberg, N., Santtila, P., Sandnabba, K., & Niemi, P. (2003). Validity of a test of

children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *Journal of Experimental Child Psychology*, *85*(1), 32–49. https://doi.org/10.1016/S0022-0965(03)00025-0

Fivush, R., & Schwarzmueller, A. (1995). Say it once again: Effects of repeated questions on children's event recall. *Journal of Traumatic Stress*, *8*(4), 555–580. https://doi.org/10.1007/BF02102889

Fritzley, V. H., & Lee, K. (2003). Do young children always say yes to yes-no questions? A metadevelopmental study of the affirmation bias. *Child Development*, *74*(5), 1297–1313.

Fritzley, V. H., Lindsay, R. C. L., & Lee, K. (2013). Young children's response tendencies toward yes-no questions concerning actions. *Child Development*, *84*(2), 711–725. https://doi.org/10.1111/cdev.12006.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical Guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1–13. https://doi.org/http://pareonline.net/getvn.asp?v=17&n=3

Garven, S., Wood, J. M., Malpass, R. S., & Shaw, J. S. (1998). More than suggestion: the effect of interviewing techniques from the McMartin Preschool case. *The Journal of Applied Psychology*, *83*(3), 347–359. https://doi.org/10.1037/0021-9010.83.3.347

Geddie, L., Fradin, S., & Beer, J. (2000). Child characteristics which impact accuracy of recall and suggestibility in preschoolers: Is age the best predictor? *Child Abuse and Neglect*, *24*(2), 223–235. https://doi.org/10.1016/S0145-2134(99)00133-7

Gelman, R., Meck, E., & Merkin, S. (1986). Young children's numerical competence. *Cognitive Development*, *1*(1), 1–29.

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Gignac, G. E., & Powell, M. B. (2009). A psychometric evaluation of the Gudjonsson Suggestibility Scales: Problems associated with measuring suggestibility as a difference score composite. *Personality and Individual Differences*, *46*(2), 88–93. https://doi.org/10.1016/j.paid.2008.09.007

Gobbo, C. (2000). Assessing the effects of misinformation on children's recall: How and when makes a difference. *Applied Cognitive Psychology*, *14*, 163–182.

Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B. M., & Rudy, L. (1991). Children's

testimony about a stressful event: improving children's reports. *Journal of Narrative and Life History*, *1*(1), 69–99.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*(1), 26–37.

Gordon, B. N., Ornstein, P. A., Nida, R. E., Follmer, A., Crenshaw, M. C., & Albert, G. (1993). Does the use of dolls facilitate children's memory of visits to the doctor? *Applied Cognitive Psychology*, *7*, 459–474.

Gudjonsson, G. H. (1983). Suggestibility, intelligence, memory recall and personality: An experimental study. *British Journal of Psychiatry*, *142*(1), 35–37. https://doi.org/10.1192/bjp.142.1.35

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, *5*(3), 303–314. https://doi.org/10.1016/0191-8869(84)90069-2

Gudjonsson, G. H. (1987a). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology*, *26*(3), 215–221.

Gudjonsson, G. H. (1987b). Historical background to suggestibility: How interrogative suggestibility differs from other types of suggestibility. *Personality and Individual Differences*, *8*(3), 347–355. https://doi.org/10.1016/0191-8869(87)90035-3

Gudjonsson, G. H. (1989). Compliance in an interrogative situation: A new scale. *Personality and Individual Differences*, *10*(5), 535–540. https://doi.org/10.1016/0191-8869(89)90035-4

Gudjonsson, G. H. (1992). Interrogative suggestibility: factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences*, *13*(4), 479–481. https://doi.org/10.1016/0191-8869(92)90077-3

Gudjonsson, G. H. (1997). *The Gudjonsson Suggestibility Scales manual*. Psychology Press Hove.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. John Wiley & Sons.

Gudjonsson, G. H., & Clare, I. C. H. (1995). The relationship between confabulation and intellectual ability, memory, interrogative suggestibility and acquiescence. *Personality and Individual Differences*, *19*(3), 333–338. https://doi.org/10.1016/0191-8869(95)00070-M

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour*, *1*, 83–104.

Gudjonsson, G. H., & Henry, L. A. (2003). Child and adult witnesses with intellectual disability: The importance of suggestibility. *Legal and Criminological Psychology*, *8*, 241–252. https://doi.org/10.1348/135532503322363013

Gudjonsson, G. H., & Sigurdsson, J. F. (1996). The relationship of confabulation to the memory, intelligence, suggestibility and personality of prison inmates. *Applied Cognitive Psychology*, *10*(1), 85–92.

Gudjonsson, G. H., & Singh, K. K. (1984). Interogative suggestibility and deliquent boys: An empirical validation study. *Personality and Individual Differences*, *5*(4), 425–430.

Gudjonsson, G. H., Vagni, M., Maiorano, T., & Pajardi, D. (2016). Age and memory related changes in children's immediate and delayed suggestibility using the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, *102*, 25–29. https://doi.org/10.1016/j.paid.2016.06.029

Gudjonsson, G. H., & Young, S. (2010). Does confabulation in memory predict suggestibility beyond IQ and memory? *Personality and Individual Differences*, *49*(1), 65–67. https://doi.org/10.1016/j.paid.2010.03.014

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures 1. *Social Indicators Research*, *45*(1), 153–171. https://doi.org/10.1023/A:1006941729637

Henry, L. A., & Gudjonsson, G. H. (2003). Eyewitness memory, suggestibility, and repeated recall sessions in children with mild and moderate intellectual disabilities. *Law and Human Behavior*, *27*(5), 481–505. https://doi.org/10.1023/A:1025434022699

Henry, L. A., & Gudjonsson, G. H. (2007). Individual and developmental differences in eyewitness recall and suggestibility in children with intellectual disabilities. *Applied Cognitive Psychology*, *21*, 361–381. https://doi.org/10.1002/acp

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447

Hünefeldt, T., Lucidi, A., Furia, A., & Rossi-Arnaud, C. (2008). Age differences in the interrogative suggestibility of children's memory: Do shift scores peak around 5-6 years of age? *Personality and Individual Differences*, *45*(6), 521–526. https://doi.org/10.1016/j.paid.2008.06.005

Imhoff, M. C., & Baker-Ward, L. (1999). Preschoolers' suggestibility: Effects of developmentally appropriate language and interviewer supportiveness. *Journal of Applied Developmental Psychology*, *20*(3), 407–429.

Kagan, J., Snidman, N., Arcus, D., & Reznick, J. S. (1994). *Galen's prophecy: Temperament in human nature.* Basic Books.

Karpinski, A. C., & Scullin, M. H. (2009). Suggestibility under pressure: Theory of mind, executive function, and suggestibility in preschoolers. *Journal of Applied Developmental Psychology*, *30*(6), 749–763. https://doi.org/10.1016/j.appdev.2009.05.004

Kline, P. (2013). *The handbook of psychological testing*. New York: Routledge.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Kulkofsky, S. (2010). The effects of verbal labels and vocabulary skill on memory and suggestibility. *Journal of Applied Developmental Psychology*, *31*(6), 460–466. https://doi.org/10.1016/j.appdev.2010.09.002

Kulkofsky, S., Wang, E., & Ceci, S. J. (2008). Do better stories make better memories? Narrative quality and memory accuracy in preschool children. *Applied Cognitive Psychology*, *22*, 21–38. https://doi.org/10.1002/acp.1326

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse and Neglect*, *31*, 1201–1231. https://doi.org/10.1016/j.chiabu.2007.03.021

Lecce, S., Cavallini, E., & Pagnin, A. (2010). *La teoria della mente nell'arco di vita*. Il Mulino.

Lecce, S., & Pagnin, A. (2007). *Il lessico psicologico: la teoria della mente nella vita quotidiana*. Il Mulino.

Lee, K. (2004). Age, neuropsychological, and social cognitive measures as predictors of individual differences in susceptibility to the misinformation effect. *Applied Cognitive Psychology*, *18*(8), 997–1019. https://doi.org/10.1002/acp.1075

Lehman, E. B., McKinley, M. J., Thompson, D. W., Leonard, A. M., Liebman, J. I., & Rothrock, D. D. (2010). Long-term stability of young children's eyewitness accuracy, suggestibility, and resistance to misinformation. *Journal of Applied Developmental Psychology*, *31*(2),

145–154. https://doi.org/10.1016/j.appdev.2009.11.007

Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology*, *31*(4), 758–758. https://doi.org/10.1037/0012-1649.31.5.758

Lepore, S. J., & Sesco, B. (1994). Distorting children's reports and interpretations of events through suggestion. *Journal of Applied Psychology*, *79*(1), 108–120. https://doi.org/http://dx.doi.org/10.1037/0021-9010.79.1.108

Lewy, J., Cyr, M., & Dion, J. (2015). Impact of interviewers' supportive comments and children's reluctance to cooperate during sexual abuse disclosure. *Child Abuse and Neglect*, *43*, 112–122. https://doi.org/10.1016/j.chiabu.2015.03.002

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school age children. *Journal of Cultural and Evolutionary Anthropology*, *4*, 231–246. https://doi.org/10.1556/JCEP.4.2006.3

Liebman, J. I., McKinley-Pace, M. J., Leonard, A. M., Sheesley, L. A., Gallant, C. L., Renkey, M. E., & Lehman, E. B. (2002). Cognitive and psychosocial correlates of adults' eyewitness accuracy and suggestibility. *Personality and Individual Differences*, *33*(1), 49–66. https://doi.org/10.1016/S0191-8869(01)00135-0

Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology. Human Learning and Memory*, *4*(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585–589.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*. https://doi.org/10.1016/S0193-953X(05)70059-9

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, *5*(1), 86–88.

Manly, T., Robertson, I., Anderson, V., & Niparko, J. (1999). The Test of Everyday Attention for Children (TEA-Ch) Thames Valley Test Company. *Bury St. Edmunds, UK*.

Mann, S. A., Vrij, A., Fisher, R. P., & Robinson, M. (2008). See no lies, hear no lies: Differences in discrimination accuracy and response bias when watching or listening to police suspect interviews. *Applied Cognitive Psychology*, *22*, 1062–1071. https://doi.org/10.1002/acp

Marzocchi, G. M., Re, A. M., & Cornoldi, C. (2010). *BIA–Batteria Italiana per l'ADHD*. Trento: Erickson.

Mastroberardino, S., & Marucci, F. S. (2013). Interrogative suggestibility: Was it just compliance or a genuine false memory? *Legal and Criminological Psychology*, *18*(2), 274–286. https://doi.org/10.1111/j.2044-8333.2012.02048.x

Mazzoni, G. (2003). *Si può credere a un testimone?: la testimonianza e le trappole della memoria*. Il mulino.

Mazzoni, G., Loftus, E. F., & Kirsch, I. (2001). Changing beliefs about implausible autobiographical events: A little plausibility goes a long way. *Journal of Experimental Psychology: Applied*, *7*(1), 51–59. https://doi.org/10.1037/1076-898X.7.1.51

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*(1), 1–16.

McFarlane, F., & Powell, M. B. (2002). The Video Suggestibility Scale for Children: How generalizable is children's performance to other measures of suggestibility? *Behavioral Sciences and the Law*, *20*(6), 699–716. https://doi.org/10.1002/bsl.492

McFarlane, F., Powell, M. B., & Dudgeon, P. (2002). An examination of the degree to which IQ, memory performance, socio-economic status and gender predict young children's suggestibility. *Legal and Criminological Psychology*, *7*, 227–239.

McGroarty, A., & Thomson, H. (2013). Negative emotional states, life adversity, and interrogative suggestibility. *Legal and Criminological Psychology*, *18*(2), 287–299. https://doi.org/10.1111/j.2044-8333.2012.02046.x

Mehrani, M. B. (2010). What is biased? Children's strategies or the structure of yes/no questions? *First Language*, *31*(2), 214–221. https://doi.org/10.1177/0142723710391886

Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle a childhood. *British Journal of Developmental Psychology*, *24*(1), 181–196. https://doi.org/10.1348/026151005X80174

Melinder, A., Endestad, T. O. R., & Magnussen, S. (2006). Relations between episodic memory, suggestibility, theory of mind, and cognitive inhibition in the preschool child. *Scandinavian*

Journal of Psychology, *47*(6), 485–495. https://doi.org/10.1111/j.1467-9450.2006.00542.x

Melinder, A., Goodman, G. S., Eilertsen, D. E., & Magnussen, S. (2004). Beliefs about child witnesses: A survey of professionals. *Psychology, Crime & Law*, *10*(4), 347–365.

Miles, K. L., Powell, M. B., & Stokes, M. A. (2004). A comparison of the effectiveness of two suggestibility paradigms in predicting prescholers' tendency to report a non-experienced event. *Applied Cognitive Psychology*, *18*(8), 1021–1036. https://doi.org/10.1002/acp.1074

Morgan, C. A., Southwick, S., Steffian, G., Hazlett, G. A., & Loftus, E. F. (2013). Misinformation can influence memory for recently experienced, highly stressful events. *International Journal of Law and Psychiatry*, *36*(1), 11–17. https://doi.org/10.1016/j.ijlp.2012.11.002

Muir-Broaddus, J., King, T., Downey, D., & Petersen, M. (1998). Conservation as a predictor of individual differences in children's susceptibility to leading questions. *Psychonomic Bulletin & Review*, *5*, 454–458. https://doi.org/http://dx.doi.org/10.3758/BF03208821

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development*, *67*(6), 3219–3237. https://doi.org/10.1111/j.1467-8624.1996.tb01910.x

Nunnally, J. C. (1978). *Psychometric theory.* (2nd ed.). New York: McGraw-Hill.

Olafson, E. (2007). Children's memory and suggestibility. *KC Faller, Interviewing Children about Sexual Abuse: Controversies and Best Practice*, 10–34.

Otgaar, H., & Candel, I. (2011). Children's false memories: different false memory paradigms reveal different results. *Psychology, Crime & Law*, *17*(6), 513–528. https://doi.org/10.1080/10683160903373392

Parisi, M., Manno, M., & Fadda, R. (2007). Teoria della mente, coinvolgimento emotivo e suggestionabilità. In I. Petruccelli, V. Verrastro, & M. Santilli (Eds.), *Memoria e suggestionabilità nell'età evolutiva. Studi e ricerche* (pp. 175–199). Milano: FrancoAngeli.

Pears, K., & Fisher, P. A. (2005). Emotion understanding and theory of mind among maltreated children in foster care: Evidence of deficits. *Development and Psychopathology*, *17*(1), 47–65. https://doi.org/10.1017/S0954579405050030

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Peters, M. J. V., Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology*, *19*(2), 213–232.

Peterson, C., & Biggs, M. (1997). Interviewing children about trauma: Problems with "specific" questions. *Journal of Traumatic Stress*, *10*(2), 279–290. https://doi.org/10.1023/A:1024882213462

Pichler, A. S., & Sharman, S. J. (2017). Police interview quality and inconsistencies raised during cross-examination: Is there a connection? In *Safer Communities, Safer Relationships*. Prato, Italy.

Poole, D. A., Dickinson, J. J., Brubacher, S. P., Liberty, A. E., & Kaake, A. M. (2014). Deficient cognitive control fuels children's exuberant false allegations. *Journal of Experimental Child Psychology*, *118*(1), 101–109. https://doi.org/10.1016/j.jecp.2013.08.013

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals.* American Psychological Association.

Poole, D. A., & Lindsay, D. S. (1995). Interviewing preschoolers: effects of nonsuggestive techniques, parental coaching, and leading questions on reports of nonexperienced events. *Journal of Experimental Child Psychology*, *60*, 129–154.

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, *23*(5), 517–537. https://doi.org/10.1023/A:1022344128649

Powell, M. B. (2008). Designing effective training programs for investigative interviewers of children. *Current Issues in Criminal Justice*, *20*(2), 189–208.

Powell, M. B., Fisher, R. P., & Wright, R. (2005). Investigative interviewing. *Psychology and Law: An Empirical Perspective*, 11–42.

Powell, M. B., Hughes-Scholes, C. H., & Sharman, S. J. (2012). Skill in Interviewing Reduces Confirmation Bias. *Journal of Investigative Psychology and Offender Profiling*, *9*(2), 126–134. https://doi.org/10.1002/jip.1357

Powell, M. B., & Snow, P. C. (2007). Recommendations for eliciting a disclosure of abuse from a young child. *Australian Police Journal*, 76–78.

Powell, M. B., Westera, N., Goodman-Delahunty, J., & Pichler, A. S. (2016). Court questioning. In *An evaluation of how evidence is elicited from complainants of child sexual abuse.* (pp. 187–195). Sydney, Australia: Royal Commission into Institutional Responses to Child Sexual Abuse.

Pozzulo, J. D., Coplan, R. J., & Wilson, J. (2005). Not too "shy" to help catch a thief: Recall memory of shy versus non-shy witnesses. *Personality and Individual Differences*, *39*(8),

1361–1372. https://doi.org/10.1016/j.paid.2005.06.010

Pozzulo, J. D., Crescini, C., Lemieux, J. M. T., & Tawfik, A. (2007). The effect of shyness on eyewitness memory and the susceptibility to misinformation. *Personality and Individual Differences*, *43*(7), 1656–1666. https://doi.org/10.1016/j.paid.2007.05.001

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, *4*, 515–526. https://doi.org/10.1016/j.celrep.2011.1011.1001.7.

Quas, J. A., & Schaaf, J. M. (2002). Children's memories of experienced and nonexperienced events following repeated interviews. *Journal of Experimental Child Psychology*, *83*(4), 304–338. https://doi.org/10.1016/S0022-0965(02)00150-9

Quas, J. A., Wallin, A. R., Papini, S., Lench, H. C., & Scullin, M. H. (2005). Suggestibility, social support, and memory for a novel experience in young children. *Journal of Experimental Child Psychology*, *91*(4), 315–341. https://doi.org/10.1016/j.jecp.2005.03.008

Raju, S., Goodman, G. S., Bidrose, S., Pipe, M.-E., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology*, *72*(4), 235–270. https://doi.org/10.1006/jecp.1999.2491

Raju, S., Qin, J., Schaaf, J. M., & Goodman, G. S. (1997). Individual differences in children's and adults' suggestibility and false event memory. *Learning and Individual Differences*, *9*(4), 359–390. https://doi.org/10.1016/S1041-6080(97)90014-5

Raven, J. C. (1998). *Raven's progressive matrices*. Oxford Psychologists Press.

Revelle, W. (2017). An introduction to the psych package: Part II Scale construction and psychometrics.

Roberts, K. P., & Powell, M. B. (2005). The relation between inhibitory control and children's eyewitness memory. *Applied Cognitive Psychology*, *19*(8), 1003–1018. https://doi.org/10.1002/acp.1141

Rocha, E. M. (2003). *Children's memories of dental procedures: effects of individual differences, question type and temporal delay*. University of Saskatchewan. https://doi.org/10.1017/CBO9781107415324.004

Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development*, *3*(4), 415–443. http://dx.doi.org/10.1207/S15327647JCD3,4-03

Roebers, C. M., Moga, N., & Schneider, W. (2001). The role of accuracy motivation on children's and adults' event recall. *Journal of Experimental Child Psychology*, *78*(4), 313–29. https://doi.org/10.1006/jecp.2000.2577

Roebers, C. M., & Schneider, W. (2001a). Individual differences in children's eyewitness recall: The influence of intelligence and shyness. *Applied Developmental Science*, *5*(1), 9–20. https://doi.org/10.1207/S1532480XADS0501

Roebers, C. M., & Schneider, W. (2001b). Individual differences in children 's eyewitness recall: the influence of intelligence and shyness. *Applied Developmental Science*, *5*(2), 9–20. https://doi.org/10.1207/S1532480XADS0501_2

Roebers, C. M., & Schneider, W. (2002). Stability and consistency of children's event recall. *Cognitive Development*, *17*(1), 1085–1103. https://doi.org/10.1016/S0885-2014(02)00075-8

Roebers, C. M., & Schneider, W. (2005a). Individual differences in young children's suggestibility: Relations to event memory, language abilities, working memory, and executive functioning. *Cognitive Development*, *20*(3), 427–447. https://doi.org/10.1016/j.cogdev.2005.05.006

Roebers, C. M., & Schneider, W. (2005b). The strategic regulation of children's memory performance and suggestibility. *Journal of Experimental Child Psychology*, *91*(1), 24–44. https://doi.org/10.1016/j.jecp.2005.01.001

Roediger, H. L., & McDermott, K. B. (1995). Create false memories: Remembering words not presented in lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(4), 803–814.

Roma, P., Sabatello, U., Verrastro, G., & Ferracuti, S. (2011). Comparison between Gudjonsson Suggestibility Scale 2 (GSS2) and Bonn Test of Statement Suggestibility (BTSS) in measuring children's interrogative suggestibility. *Personality and Individual Differences*, *51*(4), 488–491. https://doi.org/10.1016/j.paid.2011.05.003

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rossi, L., Benatti, F., Pesce, S., Oppo, A., & Avato, F. M. (2011). Suggestionabilità interrogativa: versione italiana del Bonn Test of Statement Suggestibility (BTSS). *Psichiatria Dell'infanzia E Dell'adolescenza*, *78*(3), 684–696.

Saywitz, K. J., & Snyder, L. (1996). Narrative elaboration: test of a new procedure for interviewing children. *Journal of Consulting and Clinical Psychology*, *64*(6), 1347–1357.

https://doi.org/10.1037/0022-006X.64.6.1347

Schaaf, J. M., Alexander, K. W., & Goodman, G. S. (2008). Children's false memory and true disclosure in the face of repeated questions. *Journal of Experimental Child Psychology*, *100*(3), 157–185. https://doi.org/10.1016/j.jecp.2007.09.002

Schacter, D. L., Kagan, J., & Leichtman, M. D. (1995). True and false memories in children and adults: a cognitive neuroscience perspective. *Psychology, Public Policy, and Law*, *1*(2), 411–428.

Schneider, W., & Sodian, B. (1991). A longitudinal study of young children's memory behavior and performance in a sort-recall task. *Journal of Experimental Child Psychology*, *51*(1), 14–29. https://doi.org/10.1016/0022-0965(91)90075-4

Schneider, W., & Weinert, F. E. (1995). Memory development during early and middle childhood: Findings from the Munich Longitudinal Study (LOGIC). *Memory Performance and Competencies: Issues in Growth and Development*, 263–279.

Schooler, J. W., & Loftus, E. F. (1986). Individual differences and experimentation: Complementary approaches to interrogative suggestibility. *Social Behaviour*, *I*(July), 105–112.

Scullin, M. H., & Bonner, K. (2006). Theory of mind, inhibitory control, and preschool-age children's suggestibility in different interviewing contexts. *Journal of Experimental Child Psychology*, *93*(2), 120–138. https://doi.org/10.1016/j.jecp.2005.09.005

Scullin, M. H., & Ceci, S. J. (2001). A suggestibility scale for children. *Personality and Individual Differences*, *30*(5), 843–856. https://doi.org/10.1016/S0191-8869(00)00077-5

Scullin, M. H., Kanaya, T., & Ceci, S. J. (2002). Measurement of individual differences in children's suggestibility across situations. *Journal of Experimental Psychology. Applied*, *8*(4), 233–246. https://doi.org/10.1037/1076-898X.8.4.233

Shapiro, L. R. (2006). The effects of question type and eyewitness temperament on accuracy and quantity of recall for a simulated misdemeanor crime. *Emporia State Research Studies*, *43*(1), 1–7.

Shapiro, L. R., Blackford, C., & Chen, C. F. (2005). Eyewitness memory for a simulated misdemeanor crime: The role of age and temperament in suggestibility. *Applied Cognitive Psychology*, *19*(3), 267–289. https://doi.org/10.1002/acp.1089

Shapiro, L. R., & Purdy, T. L. (2005). Suggestibility and source monitoring errors: Blame the interview style, interviewer consistency, and the child's personality. *Applied Cognitive*

Psychology, *19*(4), 489–506. https://doi.org/10.1002/acp.1093

Siegal, M., Waters, L. J., & Dinwiddy, L. S. (1988). Misleading children: Causal attributions for inconsistency under repeated questioning. *Journal of Experimental Child Psychology*, *45*(3), 438–456. https://doi.org/10.1016/0022-0965(88)90041-0

Singh, K. K., & Gudjonsson, G. H. (1987). The internal consistency of the "shift" factor on the Gudjonsson suggestibility scale. *Personality and Individual Differences*, *8*(2), 265–266. https://doi.org/10.1016/0191-8869(87)90183-8

Singh, K. K., & Gudjonsson, G. H. (1992). Interrogative suggestibility among adolescent boys and its relationship with intelligence, memory, and cognitive set. *Journal of Adolescence*, *15*(2), 155–161.

Smith, K., & Gudjonsson, G. H. (1986). Investigation of the responses of fakers and non-fakers on the Gudjonsson Suggestibility Scale (GSS). *Medicine, Science, and the Law*, *26*(1), 66–71. https://doi.org/10.3868/s050-004-015-0003-8

Snow, P. C., Powell, M. B., & Sanger, D. D. (2012a). Clinical forum language and communication disorders in adolescents oral language competence, young speakers, and the law. *Language, Speech, and Hearing Services in Schools*, *43*(October), 496–506. https://doi.org/10.1044/0161-1461(2012/11-0065)

Snow, P. C., Powell, M. B., & Sanger, D. D. (2012b). Oral language competence, young speakers, and the law. *Clinical Forum Language and Communication Disorders in Adolescents Oral Language Competence, Young Speakers, and the Law*, *43*(October), 496–506. https://doi.org/10.1044/0161-1461(2012/11-0065)

Sternberg, K. J., Lamb, M. E., Davies, G. M., & Westcott, H. L. (2001). The Memorandum of Good Practice: theory versus application. *Child Abuse & Neglect*, *25*(5), 669–81. https://doi.org/http://dx.doi.org/10.1016/S0145-2134(01)00232-0

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Esplin, P. W., Redlich, A., & Sunshine, N. (1996). The relation between investigative utterance types and the informativeness of child witnesses. *Journal of Applied Developmental Psychology*, *17*(3), 439–451. https://doi.org/10.1016/S0193-3973(96)90036-2

Stolzenberg, S. N., & Lyon, T. D. (2014). Abuse and disclosure in criminal trials, *20*(1), 19–30. https://doi.org/10.1037/a0035000

Templeton, L. M., & Wilcox, S. A. (2000). A tale of two representations: the misinformation effect and children's developing theory of mind. *Child Development*, *71*(2), 402–416. https://doi.org/10.1111/1467-8624.00153

Thomsen, Y., & Berntsen, D. (2005). Knowing that I didn't know: Preschoolers' understanding of their own false belief is a predictor of assents to fictitious events. *Applied Cognitive Psychology*, *19*(4), 507–527. https://doi.org/10.1002/acp.1123

Vagni, M., Maiorano, T., Pajardi, D., & Gudjonsson, G. H. (2015). Immediate and delayed suggestibility among suspected child victims of sexual abuse. *Personality and Individual Differences*, *79*, 129–133. https://doi.org/10.1016/j.paid.2015.02.007

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science*, *7*(3), 261–269. https://doi.org/10.1111/j.1467-7687.2004.00344.x

Volpini, L., Melis, M., Petralia, S., & Rosenberg, M. D. (2016). Measuring children's suggestibility in forensic interviews. *Journal of Forensic Sciences*, *61*(1), 104–108. https://doi.org/10.1111/1556-4029.12987

Wakefield, H. (2006). Guidelines on investigatory interviewing of children: What is the consensus in the scientific community? *American Journal of Forensic Psychology*, *24*(3), 57.

Walker, A. G. (2013). *Handbook on questioning children. A linguistic perspective.* (3rd ed.). Washington DC: ABA Center on Children and the Law.

Ward, R. A., & Loftus, E. F. (1985). Eyewitness performance in different psychological types. *The Journal of General Psychology*, *112*(2), 191–200.

Warren, A. R., Hulse-Trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility in children. *Law and Human Behavior*, *15*(3), 273–285. https://doi.org/10.1007/BF01061713

Waterman, A. H., Blades, M., & Spencer, C. (2001). Interviewing children and adults: The effect of question format on the tendency to speculate. *Applied Cognitive Psychology*, *15*(5), 521–531. https://doi.org/10.1002/acp.741

Waterman, A. H., Blades, M., & Spencer, C. (2004). The effect of question format and interviewer knowledge on children's "don't know" responses. *British Journal of Developmental Psychology*, *22*, 335–348. https://doi.org/10.1348/0261510041552710

Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children: Manual*. Psychological Corporation.

Welch-Ross, M. K. (1999a). Interviewer knowledge and preschoolers' reasoning about knowledge states moderate suggestibility. *Cognitive Development*, *14*(3), 423–442. https://doi.org/10.1016/S0885-2014(99)00013-1

Welch-Ross, M. K. (1999b). Preschoolers' understanding of mind: Implications for suggestibility.

Cognitive Development, *14*, 101–131. https://doi.org/10.1016/S0885-2014(99)80020-3

Welch-Ross, M. K., Diecidue, K., & Miller, S. A. (1997). Young children's understanding of conflicting mental representation predicts suggestibility. *Developmental Psychology*, *33*(1), 43–53. https://doi.org/10.1037/0012-1649.33.1.43

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wohlwill, J. F. (1973). *The study of behavioral development*. Academic Press.

Young, K., Powell, M. B., & Dudgeon, P. (2003). Individual differences in children's suggestibility: A comparison between intellectually disabled and mainstream samples. *Personality and Individual Differences*, *35*(1), 31–49. https://doi.org/10.1016/S0191-8869(02)00138-1

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica*, *115*(2–3), 167–183. https://doi.org/10.1016/j.actpsy.2003.12.005

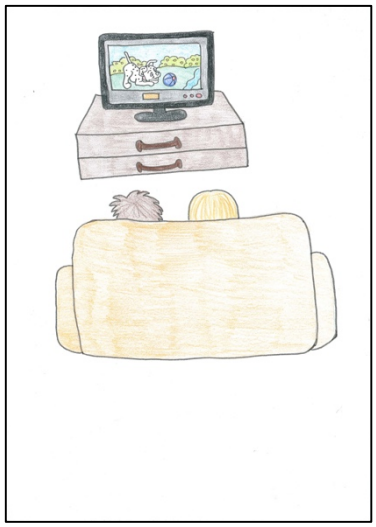
Zhu, B., Chen, C. C., Loftus, E. F., Lin, C., He, Q., Li, H., … Dong, Q. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences*, *48*(8), 889–894. https://doi.org/10.1016/j.paid.2010.02.016

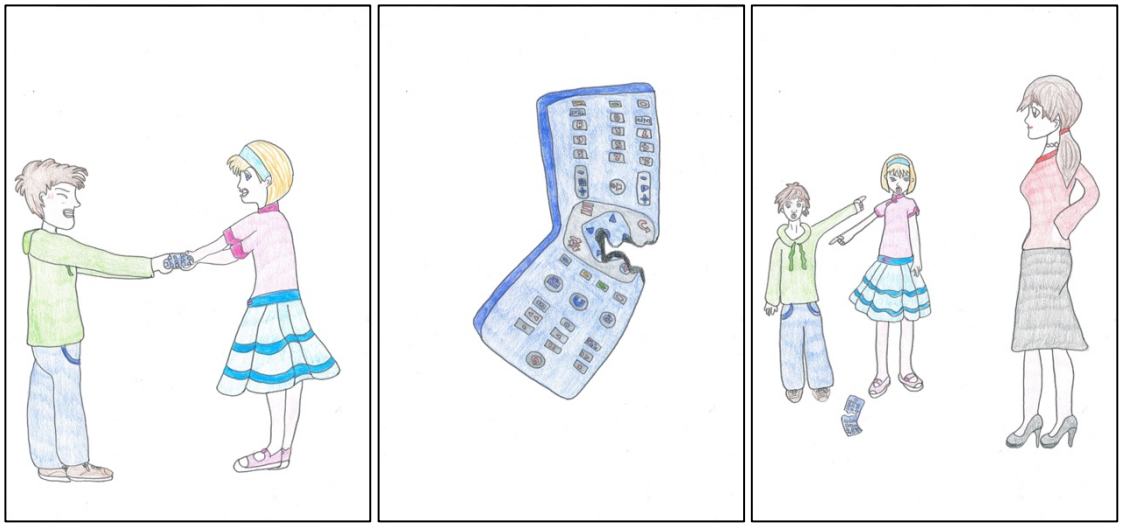
The Bicocca Suggestibility Scales (Italian version)

FORMA A

Questa è Anna. / Questo è Marco. /
Loro sono fratelli. / Marco / è più vecchio / ma Anna / è più alta /. Dopo / aver fatto i compiti / guardano insieme i cartoni. / Finiti i cartoni / inizia un film / poliziesco. / Anna vuole vedere il film / ma a Marco non piace. / Così iniziano a lottare / e il telecomando cade per terra / e si rompe. / La mamma / li sente litigare / ed entra nella stanza. / Marco / raccoglie il telecomando / e lo nasconde / dietro la schiena. / La mamma chiede: "perché litigate?". / Anna risponde: / "non litigavamo! / stavamo giocando". / Poi la mamma rivolta a Marco chiede: / "che cosa nascondi? / e lui le mostra il telecomando rotto. / La mamma si arrabbia / e vuole sapere chi lo ha rotto / Anna dice: / Marco lo ha lanciato / contro il muro. / Marco dice che non è vero. / Anna lo ha lanciato / contro la porta. / La mamma dice che per punizione / Marco / non avrebbe più giocato / con la play-station / per tutta la settimana. / Poi la mamma dice che / il telecomando non si può aggiustare, / papà / ne dovrà comprare uno nuovo. /







BISS-A32 – PRIMA SOMMINISTRAZIONE

1 La mamma si arrabbia molto quando vede il telecomando rotto?

2 Marco lancia il telecomando contro il muro?

3 Quando entra la mamma, Anna nasconde il telecomando sotto il cuscino?

4 Il telecomando era stato comprato dal papà?

5 I bambini avevano mangiato un panino con la marmellata o la Nutella?

6 Marco dice che Anna ha lanciato il telecomando contro la porta o contro il muro?

7 Marco e Anna avevano fatto merenda?

8 La mamma mette in punizione Marco o Anna?

9 La mamma dà una sberla a Anna o Marco?

10 Il papà si è arrabbiato molto quando ha saputo che si è rotto il telecomando?

11 La punizione data dalla mamma è di non guardare più la tele o giocare con la Play Station?

12 Il telecomando cade a terra durante la lotta?

13 La mamma dice che il telecomando è rotto o che si può riparare?

14 Chi dice alla mamma che non stavano litigando ma giocando, Marco o Anna?

15 Marco vuole vedere un film di fantascienza o un programma di musica?

16 Chi andrà a comprare il telecomando nuovo, la mamma o il papà?

17 Marco spegne la tele perché vuole giocare con la Play Station?

18 Chi ha rotto il telecomando la scorsa settimana, Marco o Anna?

19 Il telecomando costa molti soldi?

20 A Marco non piace il film poliziesco?

21 La mamma stava cucinando o stirando?

22 Il papà era in cucina a leggere?
23 I bambini guardano un programma di avventura o i cartoni animati?
24 Anna e Marco guardano i cartoni dopo aver fatto i compiti?
25 Marco e Anna fanno i compiti in cucina?
26 Marco vuole fare il musicista o il calciatore da grande?
27 Anna è più brava di Marco a scuola?
28 La tele è vicina alla porta o alla finestra?
29 Anna vuole fare la poliziotta da grande?
30 I due bambini guardano la TV tutti i pomeriggi o solo qualche volta?
31 La mamma era fuori a fare la spesa quando i bambini hanno iniziato a litigare per la tele?
32 Anna vuole vedere un film poliziesco o un documentario sugli animali?

BISS-A32 – DOMANDE RIPETUTE

1. La mamma non si arrabbia molto quando vede il telecomando rotto, vero?
2. Marco non lancia il telecomando contro il muro, vero?
3. Quando entra la mamma, Anna non nasconde il telecomando sotto il cuscino, vero?
4. Il telecomando era stato comprato dal papà, vero?
5. I bambini avevano mangiato un panino con la Nutella o la marmellata?
6. Marco dice che Anna ha lanciato il telecomando contro il muro o contro la porta?
7. Marco e Anna avevano fatto merenda, vero?
8. La mamma mette in punizione Anna o Marco?
9. La mamma dà una sberla a Marco o ad Anna?
10. Il papà si è arrabbiato quando ha saputo che si è rotto il telecomando, vero?
11. La punizione data dalla mamma è di non giocare con la Play Station o di non guardare più la tele?
12. Il telecomando non è caduto a terra durante la lotta ma dopo, vero?
13. La mamma dice che il telecomando si può riparare o che è rotto?
14. Chi dice alla mamma che non stavano litigando ma giocando, Anna o Marco?
15. Marco vuole vedere un programma di musica o un film di fantascienza?
16. Chi andrà a comprare il telecomando nuovo, il papà o la mamma?
17. Marco non spegne la tele perché vuole giocare con la Play Station, vero?
18. Chi ha rotto il telecomando la scorsa settimana, Marco o Anna?
19. Il telecomando costa molti soldi, vero?
20. A Marco piace molto il film poliziesco, vero?
21. La mamma stava stirando o cucinando?

22. Il papà era in cucina a leggere, vero?
23. I bambini guardano i cartoni animati o un programma di avventura?
24. Anna e Marco guardano i cartoni, prima di aver fatto i compiti, vero?
25. Marco e Anna fanno sempre i compiti in cucina, vero?
26. Marco vuole fare il calciatore o il musicista da grande?
27. Anna non è più brava di Marco a scuola, vero?
28. La tele è vicina alla finestra o alla porta?
29. Anna non vuole fare la poliziotta da grande, vero?
30. Non si sa se i due bambini guardano la TV solo qualche volta o tutti i pomeriggi?
31. La mamma non era fuori a fare la spesa quando i bambini hanno iniziato a litigare per la tele, vero?
32. Anna vuole vedere un documentario sugli animali o un film poliziesco?

BISS-A22 – PRIMA SOMMINISTRAZIONE

1. Anna vuole vedere un film poliziesco o un documentario sugli animali?
2. I bambini avevano mangiato un panino con la marmellata o la Nutella?
3. La mamma dà una sberla a Anna o Marco?
4. Il papà si è arrabbiato molto quando ha saputo che si è rotto il telecomando?
5. La punizione data dalla mamma è di non guardare più la tele o di non giocare con la Play Station?
6. Il telecomando cade a terra durante la lotta?
7. La mamma dice che il telecomando è rotto o che si può riparare?
8. Chi dice alla mamma che non stavano litigando ma giocando, Marco o Anna?
9. Marco vuole vedere un film di fantascienza o un programma di musica?
10. Chi ha rotto il telecomando la scorsa settimana, Marco o Anna?
11. Il telecomando costa molti soldi?
12. A Marco non piace il film poliziesco?
13. La mamma stava cucinando o stirando?
14. Il papà era in cucina a leggere?
15. I bambini guardano un programma di avventura o i cartoni animati?
16. Anna e Marco guardano la tele dopo aver fatto i compiti?
17. Marco e Anna di solito fanno i compiti in cucina?
18. Marco vuole fare il musicista o il calciatore da grande?
19. Anna a scuola è più brava di Marco?
20. La tele è vicina alla porta o alla finestra?

21. Anna vuole fare la poliziotta da grande?
22. I due bambini guardano la TV tutti i pomeriggi o solo qualche volta?

BISS-A22 – DOMANDE RIPETUTE

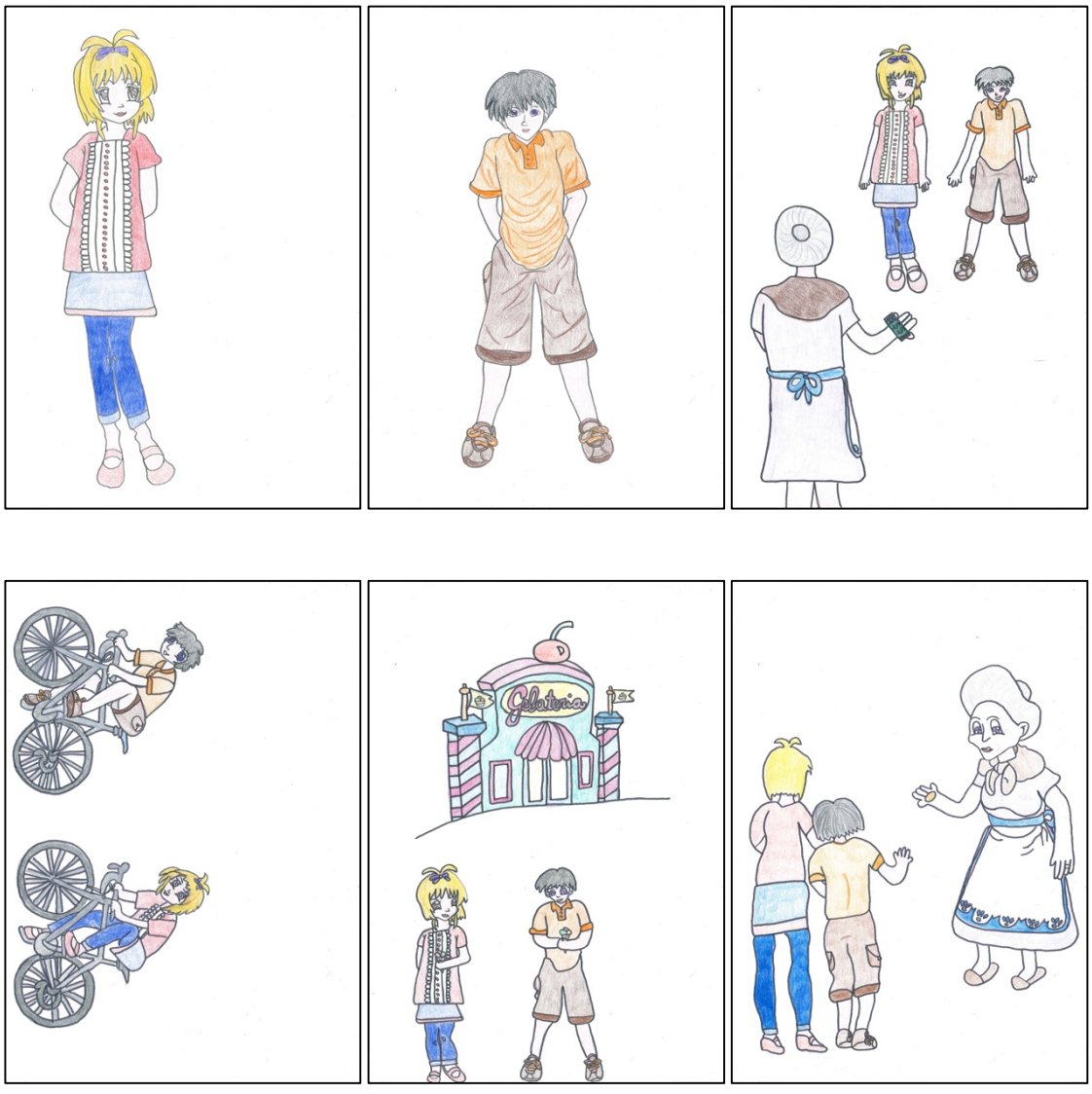
1. Anna vuole vedere un documentario sugli animali o un telefilm poliziesco?
2. I bambini avevano mangiato un panino con la Nutella o la marmellata?
3. La mamma dà una sberla a Marco o ad Anna?
4. Il papà si è arrabbiato quando ha saputo che si è rotto il telecomando, vero?
5. La punizione data dalla mamma è di non giocare con la Play Station o di non guardare più la tele?
6. Il telecomando non è caduto a terra durante la lotta ma dopo, vero?
7. La mamma dice che il telecomando si può riparare o che è rotto?
8. Chi dice alla mamma che non stavano litigando ma giocando, Anna o Marco?
9. Marco vuole vedere un programma di musica o un film di fantascienza?
10. Chi ha rotto il telecomando la scorsa settimana, Marco o Anna?
11. Il telecomando costa molti soldi, vero?
12. A Marco piace molto il telefilm poliziesco, vero?
13. La mamma stava stirando o cucinando?
14. Il papà era in cucina a leggere, vero?
15. I bambini guardano i cartoni animati o un programma di avventura?
16. Anna e Marco guardano i cartoni, prima di aver fatto i compiti, vero?
17. Marco e Anna fanno sempre i compiti in cucina, vero?
18. Marco vuole fare il calciatore o il musicista da grande?
19. Anna non è più brava di Marco a scuola, vero?
20. La tele è vicina alla finestra o alla porta?
21. Anna non vuole fare la poliziotta da grande, vero?
22. Non si sa se i due bambini guardano la TV solo qualche volta o tutti i pomeriggi?

FORMA B

Questa è Alice. / Questo è Matteo.

Loro sono cugini. / Alice / ha i capelli biondi / ma Matteo / ha gli occhi azzurri. / Loro sono in montagna / dalla nonna. / La nonna / vuole preparare una torta / ma ha finito le uova. / La nonna dice ai ragazzi: / di andare / in paese / a comprare delle uova. / E la nonna raccomanda loro / di non andare in bicicletta / perché le uova potrebbero rompersi". / La nonna dà loro i soldi. / I

bambini disobbedendo alla nonna / prendono la bici. / Arrivati in paese/ vedono una gelateria / e si comprano il gelato / anche se sanno bene / che non si può mangiare il gelato / prima di pranzo. / Arrivati in negozio / però non hanno più abbastanza soldi / e comprano solo un uovo. / Tornati a casa / danno l'uovo alla nonna. / La nonna chiede loro perché / hanno comprato un uovo soltanto /e aggiunge con tristezza: / "non posso farvi la torta"./ Alice racconta: / "Matteo ha voluto prendere la bici, / è caduto / e le ha rotte!"./ Matteo si arrabbia / e dice alla nonna: "Non è vero!" / Alice ha perso i soldi!"./ La nonna risponde con dolcezza:/ "Non credo a nessuno dei due. / Voi due mi nascondete qualcosa / ed io scoprirò cosa avete combinato!".



BISS-B32 – PRIMA SOMMINISTRAZIONE

1 Alice e Matteo sono a casa della nonna?
2 La nonna dice a Matteo che le servono 10 uova?
3 La nonna dice ai ragazzi di comprarsi un gelato?
4 Prima di arrivare in paese i bambini devono pedalare lungo una salita?
5 È estate o primavera?
6 In negozio comprano un uovo o viene loro regalato dal negoziante?
7 Prima di tornare a casa i bambini si fermano a giocare nel bosco?
8 I bambini mangiano il gelato prima o dopo il pranzo?
9 Mentre i ragazzi erano in paese, la nonna ha lavato i panni o ha rifatto i letti?
10 Alice mangia un gelato alla fragola?
11 Il gelato lo mangia solo Alice o entrambi?
12 Matteo e Alice sono cugini?
13 La casa della nonna è in montagna o in campagna?
14 Matteo dice di aver perso lui i soldi o che li ha persi Alice?
15 Il cane della nonna si chiama Tobia o Bobby?
16 La gelateria è vicino alla casa della nonna o in paese?
17 La nonna dice ai ragazzi di andare in paese in bici per far prima?
18 La nonna vuole fare la torta al cioccolato o alla marmellata?
19 Quando arrivano in paese Matteo e Alice incontrano altri loro amici?
20 Alice dice alla nonna che Matteo ha rotto le uova cadendo dalla bici?
21 La nonna ha in casa un gatto o un pappagallo?
22 Le mamme di Matteo e Alice sono sorelle?
23 La nonna punisce Matteo o non crede a nessuno dei due?
24 Matteo e Alice mangiano il gelato quando arrivano in paese?
25 Matteo ha un fratello più piccolo?
26 C'è il sole o minaccia di piovere?
27 A Matteo il gelato fa venire il mal di pancia?
28 Alice convince Matteo a comprare il gelato o lo compra di nascosto per tutti e due?
29 Quando i bambini tornano il nonno stava leggendo il giornale?
30 Visto che non ha abbastanza uova la nonna decide di fare dei biscotti o delle frittelle?
31 Matteo parte da casa da solo ed Alice lo raggiunge lungo la strada?
32 Matteo si arrabbia con Alice o con la nonna?

BISS-B32 – DOMANDE RIPETUTE

1. Alice e Matteo non sono a casa della nonna, vero?
2. La nonna non dice a Matteo che le servono 10 uova, vero?
3. La nonna non dice ai ragazzi di comprarsi un gelato, vero?
4. Prima di arrivare in paese i bambini devono pedalare lungo una salita, vero?
5. È primavera o estate?
6. In negozio viene loro regalato un uovo dal negoziante o lo comprano loro?
7. Prima di tornare a casa i bambini si fermano a giocare nel bosco, vero?
8. I bambini mangiano il gelato dopo il pranzo o prima?
9. Mentre i ragazzi erano in paese, la nonna ha rifatto i letti o ha lavato i panni?
10. Alice mangia un gelato alla fragola, vero?
11. Il gelato lo mangiano entrambi o solo Alice?
12. Matteo e Alice non sono cugini, vero?
13. La casa della nonna è in campagna o in montagna?
14. Matteo dice che Alice ha perso i soldi o che li ha persi lui?
15. Il cane della nonna si chiama Bobby o Tobia?
16. La gelateria è in paese o vicino alla casa della nonna?
17. La nonna non dice ai ragazzi di andare in paese in bici per far prima, vero?
18. La nonna vuole fare la torta alla marmellata o al cioccolato?
19. Quando arrivano in paese Matteo e Alice incontrano altri loro amici, vero?
20. Alice non dice alla nonna che Matteo ha rotto le uova cadendo dalla bici, vero?
21. La nonna ha in casa un pappagallo o un gatto?
22. Le mamme di Matteo e Alice sono sorelle, vero?
23. La nonna non crede a nessuno dei due o punisce Matteo?
24. Matteo e Alice non mangiano il gelato quando arrivano in paese ma prima, vero?
25. Matteo ha un fratello più piccolo, vero?
26. Minaccia di piovere o c'è il sole?
27. A Matteo il gelato fa venire il mal di pancia, vero?
28. Alice compra il gelato di nascosto per tutti e due o convince Matteo a comprarlo?
29. Quando i bambini tornano il nonno stava leggendo il giornale, vero?
30. Visto che non ha abbastanza uova la nonna decide di fare delle frittelle o dei biscotti?
31. Matteo non va in paese da solo ma parte con Alice, vero?
32. Matteo si arrabbia con la nonna o con Alice?

BISS-B22 – PRIMA SOMMINISTRAZIONE

1. Alice e Matteo sono a casa della nonna?
2. Prima di arrivare in paese i bambini devono pedalare lungo una salita?
3. È estate o primavera?
4. In negozio comprano un uovo o viene loro regalato dal negoziante?
5. Mentre i ragazzi erano in paese, la nonna ha lavato i panni o ha rifatto i letti?
6. Alice mangia un gelato alla fragola?
7. Il gelato lo mangia solo Alice o entrambi?
8. La casa della nonna è in montagna o in campagna?
9. Matteo dice di aver perso lui i soldi o che li ha persi Alice?
10. Il cane della nonna si chiama Tobia o Bobby?
11. La nonna vuole fare la torta al cioccolato o alla marmellata?
12. Alice dice alla nonna che Matteo ha rotto le uova cadendo dalla bici?
13. La nonna ha in casa un gatto o un pappagallo?
14. Le mamme di Matteo e Alice sono sorelle?
15. La nonna punisce Matteo o non crede a nessuno dei due?
16. Matteo e Alice mangiano il gelato quando arrivano in paese?
17. Matteo ha un fratello più piccolo?
18. C'è il sole o minaccia di piovere?
19. A Matteo il gelato fa venire il mal di pancia?
20. Alice convince Matteo a comprare il gelato o lo compra di nascosto per tutti e due?
21. Quando i bambini tornano il nonno stava leggendo il giornale?
22. Visto che non ha abbastanza uova la nonna decide di fare dei biscotti o delle frittelle?

BISS-B32 – DOMANDE RIPETUTE

1. Alice e Matteo non sono a casa della nonna, vero?
2. Prima di arrivare in paese i bambini devono pedalare lungo una salita, vero?
3. È primavera o estate?
4. In negozio viene loro regalato un uovo dal negoziante o lo comprano loro?
5. Mentre i ragazzi erano in paese, la nonna ha rifatto i letti o ha lavato i panni?
6. Alice mangia un gelato alla fragola, vero?
7. Il gelato lo mangiano entrambi o solo Alice?
8. La casa della nonna è in campagna o in montagna?
9. Matteo dice che Alice ha perso i soldi o che li ha persi lui?
10. Il cane della nonna si chiama Bobby o Tobia?

11. La nonna vuole fare la torta alla marmellata o al cioccolato?
12. Alice non dice alla nonna che Matteo ha rotto le uova cadendo dalla bici, vero?
13. La nonna ha in casa un pappagallo o un gatto?
14. Le mamme di Matteo e Alice sono sorelle, vero?
15. La nonna non crede a nessuno dei due o punisce Matteo?
16. Matteo e Alice non mangiano il gelato quando arrivano in paese ma prima, vero?
17. Matteo ha un fratello più piccolo, vero?
18. Minaccia di piovere o c'è il sole?
19. A Matteo il gelato fa venire il mal di pancia, vero?
20. Alice compra il gelato di nascosto per tutti e due o convince Matteo a comprarlo?
21. Quando i bambini tornano il nonno stava leggendo il giornale, vero?
22. Visto che non ha abbastanza uova la nonna decide di fare delle frittelle o dei biscotti?

Questioning protocol about the video

1. Era aprile o maggio?
2. La signora Rossi era in vacanza?
3. La donna si chiamava Maria Rossi o Chiara Rossi?
4. La signora Rossi si trovava a Milano?
5. Pensaci bene: la signora Rossi si trovava a Milano?
6. Nella borsa c'erano dei soldi?
7. L'aggressione è avvenuta nella stazione della metropolitana o sull'autobus?
8. Uno degli aggressori ha colpito la donna con un calcio?
9. Pensaci ancora: uno degli aggressori ha colpito la donna con un calcio?
10. La signora aveva altri soldi in tasca?
11. Le grida della donna hanno spaventato gli aggressori?
12. L'orologio della donna si è rotto quando è caduta?
13. La gonna della signora si è rotta durante la rapina?
14. Te lo chiedo ancora: la gonna della signora si è rotta durante la rapina?
15. La donna è stata accompagnata in ospedale?
16. La signora si è rotta la gamba destra o sinistra?
17. Prova a pensarci ancora. La signora si è rotta la gamba sinistra o quella destra
18. Dovrà portare il gesso per un mese o due?
19. La donna aveva uno o due figli?
20. Sei sicuro? La donna aveva due figli o uno solo?
21. Tu quanti aggressori hai visto?
22. Prima di aggredire la signora Rossi, i ladri avevano derubato un'altra donna o un uomo?
23. Gli aggressori erano magri o grassi?
24. Gli aggressori erano sui 20 anni o sui 40?
25. Gli aggressori hanno minacciato la donna con un coltello o con una pistola? 26. che nazionalità erano gli aggressori?
27. Sei sicuro? Prova a pensarci bene!
28. Il passante che ha aiutato la signora Rossi era un turista?
29. Il passante si chiamava Luca o Paolo?
30. Te lo chiedo ancora: il passante si chiamava Paolo o Luca?

31. La polizia è arrivata dopo un'ora?

32. Il marito della donna è andato a prenderla in ospedale o alla centrale di Polizia?

33. La borsa della donna è stata ritrovata?

34. Sei sicuro? La borsa della donna è stata ritrovata?

35. La borsa della donna si è danneggiata nella lotta?

36. Sei sicuro? La borsa della donna si è danneggiata nella lotta?

37. Gli aggressori sono stati condannati un mese dopo il loro arresto? 38. Quanto tempo dovranno restare in galera gli aggressori?

ACKNOWLEDGMENTS

There are many amazing people I would like to thank who have been so important to me during the preparation of this thesis.

First, I would like to thank my supervisor, Professor Claudia Caprin: she has been the one introducing me to the topic of interrogative suggestibility. She granted me access to the data collected by her students and created the Bicocca Suggestibility Scales: without it, this thesis could not have existed. I also thank her for the teachings she has given me over the years.

I am very grateful to the teachers, children, and parents of all the schools involved in this project.

Thanks also to all the students and colleagues who administered the BISS to children and helped me in collecting the data for this thesis. Thanks also to Elisa Giustarini who helped with checking the accuracy of the data inserted. I could not have completed this project alone – thank you all for your hard work throughout these years.

Moreover, I feel grateful to have been granted the opportunity to work with Professor Martine Powell, who provided me with an insight into her immense knowledge and expertise, sharing precious feedbacks. Her fairness and dedication to her work are amazing and are to be admired.

My sincerest respect and gratitude go to Professor Lina Ricciardelli and her immense generosity, she provided me with motivation, patience, and endless support. She has been a model on how to cope with problems. Her constant presence and her sincere interest have been for me essential.

Thanks to Professor Nicoletta Salerni for her patience, her helpful writing feedback, and suggestions.

Thanks are also due to Professor Steve Bowe, Professor Germano Rossi and Dr. Alessio Facchin for their statistical advice and guidance.

Thanks to my fellow students at Milano-Bicocca University for their everyday presence and reassuring friendship, for the support and for having shared the same path.

I will always be grateful to Carolina and Anne Sophie for their listening and understanding, their emotional support and some editing advice also. They pushed me in persisting through the final stages.

Thanks to Bronwen, Mohammed, Hamida and Cate, who welcomed me into their research group and have been my daily company in the last year of my PhD program.

Thanks to Marta and Claudia: I know I can always count on their friendship and support.

I want to express my deepest thanks to my family, who have loved, supported and encouraged me for my entire life. Their confidence in my abilities let me pursue my goals. They have been supportive even when my decisions were not easy to accept. A special regard goes to my sister Anna, who provided me endless support, helpful tips and editing advice too.

Finally, I am grateful to Fabio, the most important person in my life. His love and support have been my energy and motivation. He has been my rock. Without his understanding, I could not have reached this goal. I love and appreciate him more than words can express.