Testifying through another tongue: 
Examining the effects of language barriers on accuracy and suggestibility in eyewitness testimonies.

Raver Gültekin

Supervisor: Torun Lindholm  
MASTER THESIS, Spring of 2018  
STOCKHOLM UNIVERSITY  
Department of Psychology
Language barriers in eyewitness testimonies may pose threats toward witnesses’ accuracy, and consequently on the outcome of judicial procedures. The present study aims to investigate the credibility and the extent of reported detail information of eyewitnesses’ testimony of a crime event, when the testimony is given in witnesses’ first language, second language, or second language through interpreter. Moreover, the study examines whether eyewitness suggestibility is affected by the language to which the testimony is provided. Participants (N=60) were exposed to a mock crime event and subsequently performed memory tests about that event. Results showed no differences in accuracy of suggestibility between experimental conditions. The personality trait social desirability showed no relation to suggestibility or the extent of inaccurate detail information provided in the present study. The findings are discussed in the context of implications, limitations and future directions.

A large body of research has investigated the psychology of eyewitnesses of crime events and the potential malleability and contaminations when producing eyewitness testimonies. Eyewitness testimonies form, to some extent, the basis for central decisions and verdicts in both the criminal investigation and the legal process, given the premise of infallible and non-existent range of forensic evidence (Lindholm, 2008; Shermer, Rose, & Hoffman, 2011; Wells & Olson, 2003). However, research has shown that eyewitness memories of crime events are fallible, and discriminating between accurate and inaccurate witness identifications is difficult (Lindholm, 2008; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015). Furthermore, Lindholm (2008) argues that there is a discrepancy between witnesses’ accuracy and the judgments on the reliability of witnesses’ verbal recall testimonies. Hence, eyewitness testimonies can lead investigators to inaccurate directions and therefore hinder criminal investigation (Albright, 2017). Moreover, inaccurate eyewitness testimonies can wrongfully convict innocent people with misidentifications (Innocence Project, 2017; Shermer, Rose, & Hoffman, 2011; Smalarz & Wells, 2015; Wells & Olson, 2003). Henceforth, from an applied forensic standpoint, acquiring accurate and rigorous eyewitness testimonies is of great importance for the legal system. Consequently, a large body of research has been dedicated to investigate in what ways eyewitness testimonies are malleable and susceptible for contaminations and to what extent (Albright, 2017; Smalarz & Wells, 2015).

Lindholm (2008), posits that stereotypic biases (based on e.g., ethnicity, ingroup/outgroup favoritism, gender) may affect the credibility assessments on eyewitness testimonies. To examine this, she conducted a study investigating the validity in judgments of in-group and out-group witnesses. Participants rated the reliability of testimonies provided by high- and low-accurate witnesses either belonging to participant’s ethnical ingroup or outgroup (Swedish or immigrant). The findings showed judgements of low reliability on outgroup (immigrants) testimonies independent of actual memory performance (high or low), suggesting biased judgments and confidence in testimonies provided by an outgroup. For judgment on ingroup
(Swedish) testimonies, participants reported higher reliability for high memory performances in contrast with low memory performances. The findings corroborate biases found in other areas where people from different groups are judged (Albright, 2017; Correll, Park, Judd, & Wittenbrink, 2002). The findings could be generalized to judgments on eyewitness testimonies provided in a language that is not the witness’ native tongue. The present study addresses this issue by investigating the effects of language barriers on the accuracy and suggestibility of eyewitness testimonies on a mock crime event.

Memory suggestibility: encoding, storage and retrieval

The human memory process is divided in three phases; encoding, storage, and retrieval (Tulving & Bower, 1974). Encoding allows registration of perception and information to memory. Storage refers to maintaining information in the memory for a period of time. Retrieval is accessing and recalling the information from memory (Tulving & Bower, 1974). Moreover, studies show that human memory is not coded or retained complete, and consequently, obstructing the capability of recalling uncontaminated and accurate memories (Loftus, 1979; Laney & Loftus, 2013). We recall events and experiences fragmentally and when there are gaps we like to fill empty spaces with knowledge and experience from other similar situations. Hannigan and Reinitz (2001) called this phenomenon schematic gap filling and inference-based memory errors. We reconstruct the course of events based on experiences and knowledge, we also attribute our inherent attitudes, stereotypes and expectations (Laney & Loftus, 2013; Loftus, Miller & Burns, 1978; Hellmann & Memon, 2016; Osborne & Davies, 2013). Furthermore, in the context of criminal investigation and legal process, confabulations and the suggestibility of memory processes can potentially cause critical ramifications (Gudjonsson, 2017).

Suggestibility is the process of memory distortion in the encoding, storage and/or retrieval process caused by deliberate or unintentional suggestion (Ceci & Bruck, 1995). Suggestions and memory distortions can occur internally (e.g., difficulties to differentiate between fantasies and imaginations from episodic memories; Hellman & Memon, 2016; Kleider et al., 2008) or externally (e.g., source monitoring or misinformation effect; Johnson, Hashtroudi, & Lindsay, 1993; Loftus, 1979; Zaragoza, Belli & Payment, 2007). Moreover, leading and suggestive questions during interrogations can cause memory distortions and consequently affect the accuracy and credibility of the witness testimony (Lamb, Sternberg & Esplin, 1994).

Retention interval refers to the period between encoding and retrieval (Shepherd, 1983). Shepherd (1983), investigated through three different experiments to what extent the memory of specific events is malleable over time. The studied retention intervals in the experiments were 1 week, 1 month or 3 months. Shepherd (1983), found strong support for increased memory contamination over time. A body of research supports the effects of retention interval on memory performance, positing high vulnerability during the period between encoding and retrieval (Ecker, Lewandowsky, Cheung, & Maybery, 2015; Gordon & Thomas, 2017).

A classic study by Loftus and Palmer (1974) demonstrated the complications of leading questions and misleading information ( misinformation) during retrieval. In the study, participants witnessed a car accident via a filmed event and a week later reported their memories of the event. By reformulating single words in questions; for example, group 1 got asked "What speed did the cars have when they bumped together?" and group 2 got asked "What speed did the cars have when they crashed together?", significant results were shown a week later when asked whether or not the participants had seen broken glass at the accident site. Group 2 responded to a greater extent that they had seen broken glass, when in fact, there was no broken
glass in the film. The findings on misinformation effect have been replicated in a vast number of studies (Laney & Loftus, 2013).

Kaplan, Van Damme, Levine, and Loftus (2016), conducted an extensive review on the literature on the relationship between emotions and vulnerability to false memories. Kaplan et al., (2016), reported a narrowing of the scope on acquired detail information when witnessing emotional intense events (such as crime events) towards features and details relevant to current intrinsic goals, that is goals congruent to individual motivation (e.g., fearful individual focuses on the source of threat and angry individual focuses on the agent obstructing a goal). The cognitive reduction of attention when engaged in emotional intense events can potentially increase the vulnerability and susceptibility to memory distortions, consequently causing overall memory deficits specifically on non-goal relevant details, such as peripheral detail information. Furthermore, Kaplan et al., (2016) posits that eyewitnesses to crime events often are exposed to extensive retention intervals between the crime event and the interrogation, consequently, increasing the possibilities of potential contaminations on the narrowed and malleable memory. Still, Kaplan et al., (2016) argues, when providing inaccurate eyewitness testimonies on emotional intense crime events, eyewitnesses conveyed genuine emotions can potentially cause ramifications due to judgments of high credibility.

**Self-regulation, ego depletion and memory**

Self-regulation encloses executive functioning processes involved in working memory, attention, volitional control, inhibition, decision-making and self-control. Harkness et al., (2015) suggest that regulating emotional and cognitive responses to an experienced violent or distressing event much likely induces ego depletion, that is a cognitive state of temporarily reduced self-regulatory capacity (Harkness et al, 2015). Consequently, Harkness et al., (2015) argue that ego depletion may function as an important mediator for the accuracy of eyewitness testimonies. When producing a testimony of a crime, a willingness to be perceived as a virtuous, cooperative and reliable eyewitness and prove one’s own memory capacity is desirable (see social desirability). Therefore, both the witnessed event and the interview situation could exhaust the cognitive resources required for accurate recall of a crime event. Furthermore, ego depletion increases the reliance on stereotype-consistent information when recollecting memories of violent or distressing event, which consequently is potentially detrimental for the ongoing investigation (Harkness et al., 2015: Kleider, Cavrak, & Knuycy, 2012; Osborne & Davies, 2013:). Gudjonsson (1997), investigated eyewitnesses’ tendency to yield to suggestions, misleading questions and the propensity in which eyewitnesses adapt their answers when given negative feedback. Gudjonsson (1997), refers to these processes as interrogative suggestibility. Otgaard, Alberts and Cuppens (2012a; 2012b), have reported a positive association between ego depletion and interrogative suggestibility. In their study, a dyad condition of high level of ego depletion contrasted with a low level of ego depletion was assessed together with a suggestibility measurement (Otgaard, Alberts and Cuppens, 2012a; 2012b). Ego depleted participants were to a greater extent susceptible to contamination than non-ego depleted participants when examined with a false memory assessment on semantically associated words, (i.e. the so –called DRM paradigm, see Roediger & McDermott, 1995).

A study by Szpitalak & Polczyk, (2014), investigated whether manipulation of ego depletion would affect participants vulnerability to misinformation. Szpitalak and Polczyk (2014), conducted a series of experiments in which participants were presented with an original event. Then, participants read a reinforcing description of the original event (experiment group was exposed to details incongruent with the original event; critical details were altered) and later either got mentally warmed-up by performing an easy task for 10 minutes or mentally fatigued
by performing a difficult task for 30 minutes. Subsequently, participants got tested on their memory about the original event. The results reported by Szpitalak and Polczyk (2014), supported reduced vulnerability to misinformation when mentally warmed-up, and that mental fatigue amplified vulnerability to misinformation.

Ego depletion is also affected by peer influence. Harkness et al. (2015), conducted an experiment where participants witnessed a crime event, followed by a cognitive depleting (experiment-group) or non-cognitive depleting (control-group) task. Later, participants in both conditions were either exposed to a confederate co-witness (inducing accurate and misleading information) or performing an individual recall memory performance test about the crime event. Harkness et al., (2015), reported an overall increased event-congruent and event-incongruent information for participants engaged in co-witness discussion. Moreover, depleted participants reported to a greater extent less event-congruent information, and a substantial increase in event-incongruent information than non-depleted participants when engaged in co-witness discussion. Suggesting increased susceptibility to peer influence as a product of inhibited self-regulation (Harkness et al., 2015).

Language barriers, memory accuracy, suggestibility and social desirability
Allison, Basquin, & Gerwing, (2017), investigated whether witness testimonies provided in English as a second language would affect the extent of produced detail information. Allison, et al., assessed the accuracy of 17 interview dyads of testimonies, one witness (non-native English speaker) and one interviewer/officer (native-English speaker), provided by participants with English as a second language and contemporaneous notes (taking notes while interviewing a witness) provided by participants with English as a native language through open-ended and cued-recall questioning. Participants who acted as witnesses watched a crime event while participants who acted as officers prepared with an unrelated task. Subsequently, the dyad of participants performed an unrelated task that served as a filler. Then, the officers interrogated the witnesses about the crime event with a structured interview protocol with open-ended questions, followed by cued-recall questions. Additionally, officers were instructed to take notes systematically. Allison, et al., (2017) assessed the interrogations with two different methods: a checklist approach (scoring provided accurate detail information) and an inductive microanalysis of face-to-face dialogue (a software analyzing transcription and linking testimony with idea units about the crime event; accurate, inaccurate, repeated accurate, or repeated inaccurate details was assessed and scored accordingly).

Participants in the study by Allison, et al., (2017) provided more accurate information with open-ended questions and reported more errors with cued-recall than with open-ended questions. When comparing methods of analysis, the face-to-face dialogue encapsulated more information (accurate and inaccurate details) than a checklist approach. Allison, et al., (2017), suggested that a face-to-face dialogue method provides a richer accuracy for eyewitness testimonies and officers notes than a checklist approach. Therefore, the authors argue, when language barriers are evident, open-ended questions facilitates the opportunities of expressions rather than cued-recall questions which in turn inhibits opportunities of expression. Consequently, inhibiting eyewitnesses’ potential detail information of a testimony.

Smith, Multhaup and Ihejirika (2017), conducted a study to investigate the effects of the misinformation effect in first and second language in an academic context. Bilingual participants with unbalanced language skills in Spanish or English initially listened to lectures in Spanish or English in various disciplines. Subsequently, participants were exposed to notes in either Spanish or English that entailed consistent, inconsistent or irrelevant details about the
lectures. Later, participants completed a forced-choice recognition task in the language of the given lecture. The results showed that when exposed to the lecture in their dominant language participants had a better recognition accuracy than when exposed to language given in the inferior language. Misinformation effect increased when exposed to cross-language stimulus, that is when participants had lectures in one language and notes in the other language.

Shaw, Garcia, and Robles, (1997) investigated cross-language post-event misinformation effects in Spanish-English bilingual witnesses in three experiments. The study used a prototypical post-event misinformation design where bilingual participants witnessed a filmed crime event, then read a post-event narrative enclosing misleading information about the crime event. Subsequently, participants either answered a forced-choice or a cued-recall questionnaire about the crime event. Experiment 1 was conducted entirely to develop and test validity in material and procedure. Experiment 2 and 3 investigated cross-language as well as same-language exposure of misinformation effect with the intent to encourage the bilingual participants to speak, write, and think in one language in the first portion of the experiment, and either the same- or cross-language in the second portion of the experiment. That is assessing the effects of induced false information in same language (Spanish-Spanish, or English-English), or in cross-language (English-Spanish, or Spanish-English) as post-event information. Memory performance was assessed with responded consistent information (correct detail), inconsistent information (incorrect detail) and no-information (absence of detail). Shaw et al., (1997), reported no significant differences in accuracy across the three experiments on the effects of post-event misinformation in cross-language conditions as in the same-language conditions. Accordingly, participants reported as robust misinformation in cross-language as in same-language condition. Moreover, confidence on provided testimony did not decrease regardless of increased contamination.

The literature and research on applied witness psychology on language barriers and the role of interpreters is scarce. Maddux (2010), conducted a review on how interpreter-mediated communication impacts the reliability and validity of forensic evaluation due to increasing linguistic diversities in our society. Consecutive- and simultaneous interpretation are two different styles of speaker-interpreter approaches. With simultaneous interpretation approach, the interpreter provides interpretation parallel to the speaker’s narrative, whereas in consecutive interpretation approach, the interpreter provides interpretation in-between pauses of the narrative (Maddux, 2010). Generally, in forensic evaluations, consecutive interpretation approach is recommended because it is less exhausting, in contrast to simultaneous interpretation, on the interpreter which subsequently inhibits possible interpretation errors (Maddux, 2010). The condition with witness-interpreter in the present study implemented a consecutive interpretation approach with direct style of interpretation (first person rendering; Maddux, 2010). Hence, research suggests that cognitive resources may be exhausted when a bilingual witness switches language across encoding and retrieval, or when providing a testimony in a second language, leading to a poorer memory and an increased susceptibility to suggestions (Hamilton, Vohs, Sellier, & Meyvis, 2011). A basic assumption is that as barriers to language increase, so will ego depletion. This increased ego depletion will result in decreased memory accuracy and increased suggestibility to false information.

The role of social desirability refers to the tendency to present oneself in socially favorable and accepted way. For example, individuals can distort responses inaccurately and misleadingly, or adjust their responses favorable to the social situation/context (Laney & Loftus, 2013). Distortion could be due to individual’s disposition (i.e. personality trait) or caused by the social situation (e.g., wanting to frame a perpetrator; Paulhus & Reid, 1991). Social desirability
corresponds to reacting, thinking or feeling in a socially approved way, high tendency to social conforming, advocating friendliness and helpfulness, and a propensity to faking good (Crowne & Marlowe, 1960; Marlowe & Crowne, 1961). Richardson & Kelly (2004), conducted a study to investigate the association between interrogative suggestibility (extent to which individuals’ yield to leading questions or shift to critical feedback during interrogation; Gudjonsson & Clark, 1986), compliance and social desirability in an adolescent forensic sample. In Richardson and Kelly (2004), participants showed stark relationship between interrogative suggestibility, compliance and social desirability when assessed with psychological test instruments. Therefore, Richardson and Kelly (2004) argues, when conducting forensic investigation on crime events, the assessments of interrogative suggestibility, compliance and social desirability are recommended. This is to prevent possible biases or wrong turns in forensic investigations. In the present study we also measure social desirability bias to see to what extent it is related to inaccurate eyewitness testimonies and interrogative suggestibility.

Schematic memory

Bartlett (1932) argued for the importance of schematic memories which can be compared to mental frameworks of reference enclosing the individual's knowledge and experiences. Bartlett (1932) posited that schematic memories facilitate the organization of information, understanding of the outside world and what can be expected from different situations. The memory of an episodic event is influenced of both the current event process and our relevant schemas surrounding that specific event. Furthermore, Bartlett suggests that these structures prompts a conflict between current events and schemas and to actively differentiating between the dyad is difficult and mentally exhausting. Alba and Hasher (1983), refers to schematic gap filling as a problem that encapsulate a risk of developing false memories (Alba & Hasher, 1983).

In line with Bartlett (1932), Tuckey and Brewer (2003) examined how schematic memories could affect witnesses’ memories of a specific crime event. In the study, participants witnessed a filmed event where two people commit a bank robbery. The event contained detail information that was either schematic-consistent (e.g., male, mask, weapons, bag, getaway car), schematic-inconsistent (e.g., robbers did not have gun, robbers did not speak, robbers did not have a getaway car) and schematic-irrelevant (e.g., robbers wearing hats, description of hats, description of the bank tellers). The participants were then interviewed about the event a short delay after exposure to stimulus, and later re-interviewed on a varying number of occasions (one, two or three additional interviews) with retention intervals ranging from three days to three months. During the interviews, the participants were assessed with both free recall and cued-recall memory tests linked to the event. The result showed that schema-irrelevant details were to the greatest extent omitted, followed by schema-inconsistent details. Furthermore, participants reported to a great extent inaccurate details (not included in the stimulus event), but which are in line with what is typical of a bank robbery schema.

Inference-based memory errors

Hannigan and Reinitz (2001) introduced the concept of causal schemas (i.e. schemas of the relationship between cause and effect) and conducted a series of experiment on how scripts on causal schemas affect the memory. In one of the experiments, images were used as stimulus material containing four different scripts; (a) eating at a restaurant, (b) getting up in the morning, (c) going grocery shopping, and (d) attending a lecture. Participants who witnessed a script with a restaurant visit saw a picture in a casual image sequence of accidentally knocking over a glass of water, as well as an effect image where someone is mopping the water from the floor. Participants who witnessed the script with the grocery store were given a causal image sequence
that showed a woman picking an orange from the bottom of a pile. The effect image showed oranges on the floor. After retention intervals ranging from 15 minutes, 24 hours or 48 hours, the participants were shown 20 slides, including both causal and effect images, and were asked to rate their recognition confidence. The authors examined false memories based on Inference-based memory errors in both directions: backward inference-based errors and forward inference-based errors. Whether causal images (e.g., "knocking water glass" or "woman pick oranges") produced memories of seeing the effect ("water being dried up" or "oranges on the floor"), so-called forward inference-based errors, and whether participants who viewed effect image (e.g., "mopping water from the floor" or "oranges on the floor") produced false memories of causal image, so-called backward conclusions.

The results in Hannigan and Reinitz (2001) study showed that participants tended to incorrectly remember causals when only effect was witnessed and the propensity for backward inference-based errors were more prone than forward inference-based errors. Participants who witnessed an effect image (oranges on the floor) to a greater extent fabricated memory of causals (woman picking orange from the bottom of the pile), than participants who witnessed the causal image had fabricated memories of the specific effect. Hannigan and Reinitz (2001) also reported that longer retention interval negatively affected participants' memory, where memory performance deteriorated over time, and that error rates increased.

A recent study on causal schemas and inference-based memory errors by Mirandola, Toffalini, Grassano, Cornoldi, & Melinder, (2014) examined whether negatively charged schematic memories causing arousal (arousal is a physiological reaction to stimuli activating attentiveness and concentration; Schachter, 1964) affected newly acquired memories. Subsequently, Mirandola et al., (2014), examined whether assessments with free recall and recognition test would affect the prevalence of memory distortion. Participants were exposed to eight various scripts all enclosing the effect of an event and with excluded cause. The events’ effects were either negative (e.g., a boy being run over by a car with blood occurring) or neutral (e.g., a boy crossing the street without an incident occurring). Participants were then assigned to either a (1) recognition test group or a (2) free recall and recognition test group. Mirandola et al., (2014) reported fewer memory distortions on negatively charged events when participants performed solely recognition test than when participants performed elaborated memory tests (free-recall and recognition test), suggesting false memories engendered by semantic elaboration as a mediating role.

The present study
The present study aims to investigate the role of language barriers to the reliability and suggestibility of eyewitnesses’ testimony of a mock crime event. The language barrier was assessed by gradually obstructing the language variable in three conditions. In the first condition participants provided witness testimony in their native language (Swedish), in the second condition participants provided witness testimony in second language (English) and in the third condition participants provided witness testimony in second language (English) through an interpreter who translated the testimony to native language (Swedish). Moreover, social desirability was assessed to investigate whether this personality trait affects the propensity of suggestibility. Suggestibility was investigated in two ways; first by misleading open question, and second by, schematic gap filling.
Predictions

i. Participants providing testimony in a second language will show a poorer memory accuracy and be more influenced by misleading questions and gap-filling schemas than participants providing testimony in a native language. Moreover, participants providing testimony in a second language through an interpreter will have the poorest and most distorted memory across all conditions.

ii. Participants high in trait social desirability and exposed to language barriers (second language; second language with interpreter) will to a greater extent be affected by suggestibility and confabulations and infer backward-inference-based errors than participants low in trait social desirability.

Method

Participants
A total of 66 students, 16 men and 50 women, between 19-62 years old (M = 26.55, SD = 7.37) were recruited to the study as witnesses from the Department of Psychology, Stockholm University. A total of six participants were excluded due to non-response in cued recall and/or recognition test, or due to participants being bilingual, henceforth, not fulfilling the validity of the present study. The remaining 60 participants were evenly distributed across the three conditions (Native language, n = 20, 19-43 years (M = 26.30, SD = 6.40); Second language, n = 20, 20-62 years (M = 29.35, SD = 9.99); Second language through interpreter, n = 20, 19-32 years (M = 24.00, SD = 3.58). In addition, a total of seven interpreters, 4 men and 3 women, participated (M age = 24.42 years, range = 19-39 years, SD = 7.11). Two male interpreters participated twice, and one female interpreter participated 12 times. Remaining interpreters (n = 4) participated once. The experiment was advertised on a notice-board at the Department of Psychology and on internet on mail lists by the IT-administrator of the department. To apply for the role of interpreter students had to fulfill the criteria of at least professional language competence in English on a 4-point scale (1= limited; 2= basic; 3= professional; 4= native language). All witnesses also reported language competence on the same 4-point scale. Participants voluntarily signed up for the experiment by sending email request to the experimenter. For participation, all psychology students could receive course credit, or a movie ticket (value 135 SEK).

Material
The stimulus used in the present study is a filmed mock crime event from the study by Lindholm, Eriksson and Memon (2007). In the event, the viewer is initially perceiving a woman entering and then sitting down in a library to study. The woman is picking up items from her bag and begins studying. A few moments later in the film, a man arrives with a book and sits at the table behind the woman. He scrolls around the book and begins reading. The woman starts looking around in her bag, then rises and walks out of the scene with the bag, together with her other belongings, left on the table. The man follows her departure with his gaze and then stands up. After this sequence, the film is immediately cut to a new scene where the man is gone, and the woman returns to her table and finds that her bag is missing. In accordance with the findings from Hannigan and Reinitz (2001) the stimulus material is filling the criteria for schematic gap filling and causal-inference-based errors and specifically backward inference effect where participants witness the effect of a theft (the missing bag) and not the cause (the man taking the bag).
Free recall

For the witnesses, the memory tests consisted of three parts. The first test was a free recall where participants were instructed to imagine that they witnessed the event in the film in reality and to recall and report a testimony as comprehensive and detailed as possible. Participants were informed to report their memory without a dialogue with the experimenter. The testimony was audio recorded with the experimenter’s telephone. In the Native language condition participants provided their testimony of the witnessed event in Swedish. In the Second language condition participants provided their testimony of the witnessed event in, English, and in the Second language through interpreter condition, participants provided their testimony of the witnessed event in English through an interpreter who translated the testimony to native language, Swedish. In the interpreter condition additional instructions were given to control for accurate witness-interpreter implementation. Participants and interpreters were briefed of the intention with the roles of witness and interpreter in the present context of witness testimonies. The witnesses were instructed with the task to report their testimony in English. The interpreters were instructed to interpret the testimony of the witnesses to the experimenter in Swedish verbatim and to mediate the testimony in a first-person narrative (direct style of interpretation). Both witness and interpreter were instructed to not dialogue or to converse. The interpreters were entitled to stop the witnesses’ narrative and ask for repetition in order to mediate interpretations correctly. Furthermore, interpreters were instructed to not elaborate on the witnesses’ recollection and testimony of the event. The witnesses were encouraged to not talk too fast and pause systematically and instructed to not correct the interpreters for the retention of language barriers. Before the free recall test, the witnesses and interpreters conducted a practice on the witness-interpreter dynamics. The practice run consisted of a crime-irrelevant schema excerpt of monologue from the movie Kill Bill Volume II (Tarantino, 2004).

Cued recall

The second memory test was a cued recall. Participants answered eight questions in writing (7 open-ended questions; 1 leading question) related to the course of event. Example of open-ended questions are, ‘Describe the appearance and clothing of the woman in the film’, and ‘Where was the man in the film sitting relative to the woman?’. One leading question was included, ‘What was the title of the book that the woman picked up from her bag?’ (this information was never shown in the film). Additionally, for each question a self-report was included on how confident participants were on their answers on a 7-point scale (1 = Not confident at all; 7 = Confident).

Recognition

The third test was a recognition test where participants were presented with eight images sequentially on a computer with the task to report for each of the images if the presented scene was included in the filmed event they had witnessed or not. Four of the images were included in the mock crime they had seen, while the remaining four images were not included in that event, so-called false images. The false images showed the title of the book the women had read (False image 1), a bookshelf, (False image 2), the man stealing the woman's bag, (False image 3), the man standing outside the library (False image 4). False image 3, is the item testing causal-inference-based errors. The images were shown one at a time, and participants answered whether they had seen each image on a 7-point scale (1 = Confident the image was not included in the event; 7 = Confident the image was included in the event).

Self-report of interrogation situation

Finally, all witnesses made self-reports on their (1) memory ability, (2) ability to provide testimony, (3) affective experience of the interrogation situation, (4) motivation to recall, (5)
nervousness and (6) judgement of a hypothetical perceiver’s judgement of their credibility (i.e. ‘How do you think a person who hears the recording with your testimony would judge your credibility?’). Moreover, witnesses in condition Second language through interpreter assessed the interpreter’s competence, cooperation and their satisfaction of the witness-interpreter dynamics. The interpreters assessed their (a) interpretation ability and (b) the ability to mediate the testimony, (c) the witness’ (?) cooperation, (d) satisfaction of witness-interpreter dynamics, (e) affective experience of the interrogation situation, (f) motivation and (g) nervousness, and (h) potential judgement on credibility of the produced witness testimony (i.e. ‘How do you think a person who hears the recording with the testimony would judge the credibility?’). In each report, all items were assessed on a 7-point scale. All memory tests, self-reports and instructions for condition Native language was translated and carried out in Swedish, and for condition Second language and Second language through interpreter, translated and carried out in English. Some instructions in both condition Second language and Second language through interpreter were communicated verbally in Swedish to control for accurate systematic performances and avoid potential missing data.

**Social desirability scale (SDS)**

Karolinska scales of personality; the social desirability scale (KSP-SD), which is a short form of the social desirability scale (SDS) by Schalling (1985), was assessed. In the scale, participants take a stand at which response best corresponds to their usual way of reacting, thinking or feeling. Example of items are ‘I have never deliberately said something that has hurt someone's feelings’; ‘I have nothing against confessing my ignorance’ and ‘If I have made a mistake, then I am willing to admit it’. Participants indicated their agreement or disagreement on a 4-point scale (1; strongly disagree; 4; strongly agree). The internal-consistency reliability (Cronbach-Alpha) for KSP-SD was and found fairly weak (10 item; α = .56).

**Procedure**

All Participants, including interpreters, were informed that all responses will be administered anonymously and that no single participant will be distinguishable in the present study. Participants were also informed about their rights to abort the experiment whenever they wanted. No participant was in advance informed about the stimulus material enclosing a crime event nor the role of witness. The present study initially consisted of two parts for the witnesses, in which witnesses first saw the witnessed event approximately one-week before the memory tests (Shepherd, 1983). However, due to unexpected difficulties in recruiting participants to such two-phase study, in combination with difficulties recruiting interpreters to this second phase for witnesses in the interpreter condition, the retention interval was excluded after consultation with the supervisor. Of the 60 participants in the study, 16 were being tested with retention interval (12 in the two conditions without interpreter; 4 participants in interpreter condition). Participants with the witness role were randomized to the three conditions. The first part, with retention interval, of the study was carried out online via qualtrics.com, a web surveys tool, where participants witnessed the filmed event on one occasion and then responded to KSP-SD. When the first part was completed, participants contacted the experimenter to book a date and time for the second part of the study which was carried out at the Department of Psychology with an approximately one-week retention interval. During the second part, witnesses performed the memory tests and self-report assessments in the following order: free recall, cued recall, recognition test and self-report. With the retention interval excluded, all participants performed the complete experiment during one occasion at the Department of Psychology with the equivalent order of exposure to stimulus event and memory testing as in the previous procedure. For condition Native language and Second language, witnesses performed the second part of the experiment individually. In condition Second language through interpreter,
witnesses and interpreter started the experiment together with the free recall. After completed free recall, the interpreters answered the questions about their view of the interpretation and could subsequently leave. The witnesses continued the experiment with following memory tests (cued recall and recognition) and self-report. When the experiment was completed, participants were debriefed on the purpose of the study and encouraged to avoid talking to others about the content of the mock crime event or the framework of the second part of the experiment. And finally, participants were thanked for their participation.

Coding
To link the witnesses’ performances anonymously, each witness attributed themselves a unique nickname to participate which they used during the experiment. To link the witnesses with the correct interpreter’s self-report assessment, the interpreter got the same nickname as code as their witness. Performance on all three memory tests was assessed by using templates from the study by Lindholm, Eriksson and Memon (2007) where the same stimulus material was used. For the free recall, memory accuracy was scored from the transcripts of audiotaped testimonies provided by the witness in condition Native language and condition Second language. For condition Second language through interpreter, accuracy was scored from the interpreted information provided by the interpreter. Across both free recall and cued recall, inaccurate details were scored separately, hence there were no withdrawal of accurate scores for each inaccurate score. Examples of accurate details for both free recall and cued recall are “red-haired woman”, “bag”, “man behind woman” and “bag gone”. Examples of inaccurate details for both free recall and cued recall are “black-haired woman”, “computer”, “long-haired man” and “man at woman’s table”. To illustrate the scoring, if a participant reported the following details either in free or cued recall: “a red-haired woman placed her bag on the table”, the participant obtained 2 points. The max sum of accurate details was 46 details in free recall, and in cued recall a max sum of 16 details. Scores were compiled and analyzed separately for free recall and cued recall. For the recognition test with 8 items, participants’ estimated confidence on a 7-point scale on whether the images were included in the filmed event or not. In addition, participants' confidence for the critical causal-inference-based error image (False image 3) was examined separately to compare estimated confidence of false schematic memory of theft across the three conditions. Each witnesses’ total social desirability test score was summarized with point distribution 1-4 for each item (10). High scores indicated high in personality trait social desirability and low scores indicate low in personality trait social desirability.

Results
Mean values on accurate and inaccurate memory performance in free recall and cued recall across conditions are presented in Table 1. Further, Table 1 shows frequency of mentioned critical causal-inference-based error detail (confabulated theft) in free recall and cued recall across conditions. Table 1 also provides estimated confidence ratings on all 8 items (4 correct images; 4 false images) in recognition test, and additionally, the critical item (False image 3) is reported separately to investigate the difference between the conditions for the causal-inference-based error image. To compare the results on memory performance across three conditions for free recall, cued recall and recognition test, an analysis of variance (ANOVA) was conducted with experimental condition (Native language; Second language; Second through interpreter) as independent variable and memory performance as dependent variable. Furthermore, to compare the differences in frequency of mentioned critical detail in free recall and cued recall across the conditions, an analysis of variance (ANOVA) was conducted with experimental condition (Native language; Second language; Second through interpreter) as
independent variable and frequency of mentioned critical detail as dependent variable. For significance criterion across all statistical tests an alpha level of .05 was used. Figure 1 is presented to convey an ocular inspection of variances between all three conditions composited memory performance summed across free and cued recall testing.

Table 1. Mean values (and standard deviations) for accurate and inaccurate details in free recall (max sum 46 accurate details) and cued recall (max sum 16 accurate details), and mean values (and standard deviations) for reported estimated confidence in cued recall (8 items). Further, and mean values (and standard deviations) for reported estimated confidence in recognition (accurate images, 4 items; false images, 3 items; critical image, 1 item). Also, frequency (and percentage) of recalled suggested theft in free recall and cued recall across conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Native language</th>
<th>Second language</th>
<th>Second language interpreter</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 20)</td>
<td>(n = 20)</td>
<td>(n = 20)</td>
<td></td>
</tr>
<tr>
<td><strong>Free recall</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accurate</td>
<td>15,35 (5,06)</td>
<td>13,25 (4,43)</td>
<td>12,95 (2,50)</td>
</tr>
<tr>
<td>Inaccurate</td>
<td>1,75 (1,41)</td>
<td>2,25 (1,51)</td>
<td>1,85 (1,26)</td>
</tr>
<tr>
<td>False schematic memory of theft&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6 (30%)</td>
<td>8 (40%)</td>
<td>6 (30%)</td>
</tr>
<tr>
<td><strong>Cued recall</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accurate</td>
<td>9,15 (3,65)</td>
<td>7,55 (3,45)</td>
<td>8,25 (3,50)</td>
</tr>
<tr>
<td>Inaccurate</td>
<td>2,35 (1,69)</td>
<td>2,55 (1,90)</td>
<td>1,35 (1,38)</td>
</tr>
<tr>
<td>Leading question&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>False schematic memory of theft&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6 (30%)</td>
<td>8 (40%)</td>
<td>7 (35%)</td>
</tr>
<tr>
<td><strong>Confidence</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accurate images</td>
<td>5,11 (1,12)</td>
<td>4,89 (1,13)</td>
<td>5,41 (1,23)</td>
</tr>
<tr>
<td>False images</td>
<td>5,60 (.92)</td>
<td>4,95 (.90)</td>
<td>5,96 (.90)</td>
</tr>
<tr>
<td>Critical image</td>
<td>6,50 (.63)</td>
<td>6,03 (1,10)</td>
<td>6,60 (.51)</td>
</tr>
<tr>
<td>Accurate images</td>
<td>3,95 (2,80)</td>
<td>5,50 (2,09)</td>
<td>5,45 (2,12)</td>
</tr>
<tr>
<td>False images</td>
<td>6,03 (1,10)</td>
<td>6,60 (1,51)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Score assessed for mentioned accurate/inaccurate detail information. For free recall a max sum of 46 accurate details and for cued recall a max sum of 16 accurate details.

<sup>b</sup> Frequency (and percentage) of proportion of participants recalling witnessing a theft.

<sup>c</sup> Confidence on cued recall answers (8 items) assessed with a 7-point scale (1 = Not confident at all; 7 = Confident).

<sup>d</sup> Estimated confidence on recognition test (total 8 items: 4 accurate images, 3 false images and 1 critical image; causal-inference-based error) assessed with a 7-point scale (1 = Confident the image was not included in the event; 7 = Confident the image was included in the event).

**Free recall**

Accurate/inaccurate detail information. The results of the ANOVA revealed no significant difference between conditions in mean number of reported accurate details, $F(2,57) = 1.98, p >$
Further, results from ANOVA showed no significant difference in the number of reported inaccurate details (e.g., wrongful descriptions of appearances, actions and items) across the conditions, $F(2,57) = .71, p > .05, \eta^2_p = .02$.

**False schematic memory of theft.** Conducted ANOVA on reported frequencies of false schematic recalls of the theft in free recall test showed no significant differences across the conditions (Native language; Second language; Second language through interpreter), $F(2,57) = .29, p > .05, \eta^2_p = .01$.

**Cued recall**

**Accurate/inaccurate detail information.** The results of the ANOVA showed no significant difference in the numbers of reported accurate details across the three conditions, $F(2,57) = 1.02, p > .05, \eta^2_p = .06$. Further, results from ANOVA showed no significant difference in mean ratings on memory performance on the range of reported inaccurate details (e.g., wrongful descriptions of appearances, actions and items) across the conditions, $F(2,57) = 2.94, p > .05, \eta^2_p = .09$. For the leading question (‘What was the title of the book that the woman picked up from her bag?’; this information was never shown in the film), no participant answered wrongfully.

**False schematic memory of theft.** Conducted ANOVA on reported frequencies of false schematic recalls of the theft in cued recall test showed no significant differences across the conditions (Native language; Second language; Second language through interpreter), $F(2,57) = .21, p > .05, \eta^2_p = .00$.

**Confidence.** The results of the ANOVA showed that there was no significant difference between participants self-reported confidence on cued recall answers across three conditions $F(2,57) = .99, p > .05, \eta^2_p = .03$.

Figure 1. Mean ratings for accurate and inaccurate detail information composited across summed free and cued recall between conditions.

**Recognition**

**Total self-reported recognition.** In condition Native language participants reported in average 82.7% correct estimated confidence for all images (8 items). In condition Second language participants reported in average 79.0% correct estimated confidence for all images. And in condition Second language through interpreter participants reported in average 87.5% correct estimated confidence for all images. An ANOVA showed that there was no significant difference between participants self-reported correct estimated confidence for all images across the three conditions $F(2,57) = 4.58, p > .05, \eta^2_p = .14$. For accurate images (4 items), ANOVA
showed no significant difference across conditions, $F(2,57) = 6.70, p > .05, \eta^2_p = .19$. Similarly, no significant difference between conditions on false images (3 items) were shown, $F(2,57) = 2.70, p > .05, \eta^2_p = .09$.

**Critical item.** For the critical causal-inference-based error image (False image 3), in condition Native language participants reported in average 56.4% correct estimated confidence. In condition Second language participants reported in average 78.6% correct estimated confidence. And in condition Second language through interpreter participants reported in average 77.8% correct estimated confidence. However, an ANOVA concluded that there is no significant difference between participants self-reported correct estimated confidence for all images across the three conditions $F(2,57) = 2.72, p > .05, \eta^2_p = .08$.

**Social desirability and memory performance**
The distribution of SDS score for all (N = 60) participants varied between 17 – 32 (M = 25.28; SD = 3.39). For condition Native language (M = 24.85; SD = 3.43), for condition Second language (M = 26.00; SD = 3.83) and for condition Second language through interpreter (M = 25.00; SD = 2.88). The correlation between personality trait SDS and the extent of inaccurate memory performances in free recall and cued recall, and estimated confidence on all false images (4 items; including critical item) in recognition was estimated with two-tailed Pearson correlation coefficient (N = 60) as displayed in Table 2.

Table 2. Correlations between SDS and inaccurate details in free recall and cued recall, and estimated confidence on all false images (4 items; including critical item) in recognition.

<table>
<thead>
<tr>
<th>SDS</th>
<th>Condition</th>
<th>Native language (n = 20)</th>
<th>Second language (n = 20)</th>
<th>Interpreter (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Free recall</td>
<td>.11</td>
<td>-.05</td>
<td>-.22</td>
</tr>
<tr>
<td></td>
<td>2. Cued recall</td>
<td>-.23</td>
<td>.27</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>3. Recognition</td>
<td>.23</td>
<td>-.20</td>
<td>-.19</td>
</tr>
</tbody>
</table>

Pearson’s $r$ showed no statistically significant correlation between SDS and reported inaccurate detail information in free recall and cued recall, and estimated confidence on all false images (4 items; including critical item) in recognition across all three conditions.

**Credibility.**
Table 3 conveys mean values (and standard deviations) of participants (witnesses and interpreters) self-reported assessment on eyewitness testimony performance and on the interrogation situation across total 6 items (total 8 items for condition: Second language through interpreter).
Table 3. Mean ratings (and standard deviations) of self-report on eyewitness testimony performance.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Native language (n = 20)</th>
<th>Second language (n = 20)</th>
<th>Interpreter (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Memory</td>
<td>4.85 (1.26)</td>
<td>4.40 (1.23)</td>
<td>5.10 (1.25)</td>
</tr>
<tr>
<td>2. Ability</td>
<td>5.45 (1.05)</td>
<td>5.65 (1.04)</td>
<td>5.40 (1.31)</td>
</tr>
<tr>
<td>3. Context</td>
<td>4.45 (1.39)</td>
<td>5.05 (1.14)</td>
<td>4.75 (1.16)</td>
</tr>
<tr>
<td>4. Motivation</td>
<td>5.80 (1.00)</td>
<td>5.40 (1.14)</td>
<td>5.70 (1.30)</td>
</tr>
<tr>
<td>5. Nervousness</td>
<td>2.05 (1.31)</td>
<td>2.05 (1.35)</td>
<td>3.35 (1.49)</td>
</tr>
<tr>
<td>6. Credibility</td>
<td>4.95 (1.43)</td>
<td>4.10 (1.02)</td>
<td>4.50 (1.46)</td>
</tr>
<tr>
<td>7. Interpreters'</td>
<td></td>
<td></td>
<td>5.95 (1.43)</td>
</tr>
<tr>
<td>8. Interpreter on</td>
<td></td>
<td></td>
<td>4.90 (1.37)</td>
</tr>
</tbody>
</table>

1 Items assessed on a 7-point scale.
2 Items only in condition Second language through interpreter.

Witnesses’ self-reported judgment on credibility of provided eyewitness testimony between all three conditions across all memory testing (free recall, cued recall and recognition) was analyzed with a multivariate analysis of variance (MANOVA) with experimental condition (Native language; Second language; Second through interpreter) as independent variable and each self-report item as dependent variables. The results of the MANOVA indicated no difference across conditions, Wilks’ Lambda = .61, F(2,57) = 2.40, p = .01, η²p = .22. Follow-up univariate ANOVAs indicated no significant difference on (1) memory ability, F(2,57) = 1.61, p > .05, η²p = .05, (2) ability to provide testimony, F(2,57) = .30, p > .05, η²p = .01, (3) affective experience of the interrogation situation, F(2,57) = 1.60, p > .05, η²p = .05. Furthermore, no indication of significant difference on (4) motivation to recall, F(2,57) = .65, p > .05, η²p = .02, (5) nervousness, F(2,57) = 5.81, p = .01, η²p = .17, and finally (6) judgement of a hypothetical perceiver’s judgement of their credibility, F(2,57) = 2.10, p > .05, η²p = .07. In condition Second language through interpreter, interpreters reported intersubjective judgement on credibility of the free recall witness testimony (‘How do you think a person who hears the recording with the testimony would judge the credibility?’). Interpreters judgment on credibility was analyzed with a two-tailed Pearson product-moment correlation coefficient to assess the relationship between interpreters judgment of credibility and the extent of provided accurate and inaccurate details. When witnesses provided a high extent of accurate detail information, interpreters judged the credibility higher than when witnesses provided a low extent of accurate detail information, r = .54, n = 20, p = .14. Further, there were no significance on judged credibility between provided high and low inaccurate detail information, r = -.13, n = 20, p = .58. However, when correlating interpreters’ judgement on credibility with summed accurate and inaccurate detail information (compiled provided accurate and inaccurate details for each witness), interpreters judged significantly higher credibility when witnesses reported a greater proportion of detail information than when witnesses reported a lesser proportion of detail information, r = .510, n = 20, p = .02. Corroborating that more detailed and
comprehensive eyewitness testimonies are judged more credible than less detailed and comprehensive eyewitness testimonies regardless of accuracy.

Discussion

Research on eyewitness testimony is ample and positing difficulties in providing accurate and credible testimonies on crime events. Overall, the results in the present study on the extent of provided accurate and inaccurate detail information between three conditions showed no significant differences across all memory testing (free recall, cued recall and recognition). However, when contrasting reported accurate and inaccurate details between Native language and Second language, tendencies of the effects of language barriers on accuracy and suggestibility could be observed but not statistically reliably shown due to lack of power caused by limitations in the present study. Furthermore, when contrasting the conditions on critical detail information (recalling witnessing a theft), results showed no variations between conditions. For the recognition test, no significant variation on confidence between conditions was shown. Moreover, the personality trait social desirability showed no significant correlations with provided inaccurate detail information across all memory testing in all three conditions.

Consequently, prediction 1: participants providing testimony in a second language will have a poorer memory accuracy and be more influenced by misleading questions and gap-filling schemas than participants providing testimony in a native language, and participants providing testimony in a second language through an interpreter will have the poorest and most distorted memory across all conditions, received no support in the present study. Moreover, prediction 2: participants high in trait social desirability and exposed to language barriers (second language; second language with interpreter) would to a greater extent than participants low in trait social desirability be affected by suggestibility and confabulations and infer backward-inference-based errors, received no support in the present study.

However, consistent and corroborating with previous findings on causal-inference-based error and specifically backward inference-based errors, when inspecting overall performances, a total of 25 (of 60) participants recalls witnessing a theft on either free recall or cued recall (or in both). Suggesting investigators in real-world criminal cases to process and act cautious with the veracity provided in detail information by eyewitnesses. This due to inhibit potential wrong turns in investigations which can stall and cause devastating ramifications (Albright, 2017; Innocence Project, 2017; Shermer, Rose, & Hoffman, 2011).

An interesting aspect in the reported inaccurate details is the substantial amount of participants wrongfully recalling (in free and cued recall) witnessing the man in the filmed event to approach the woman's table and / or the man is at the woman's table. In addition, participants mentioned inaccurately that the man is looking in the woman's bag or at her computer (inaccurate item) without stealing any item. These are specific and frequent descriptions of suspicions towards the man in the film. The filmed event is in addition enclosing a substantial amount of crime irrelevant events (such as the woman fumbling with / and searching for her belongings and items; camera angle pointing towards corridor; woman placing and replacing items on the table). These crime irrelevant events, along with the participants judgement of suspicion towards the man’s behavior (man peeking at the woman and her belongings), can potentially cause disturbances in the fragmental memory. Consequently, infer inaccurate details in the fragmental crime event. Corroborating previous research, that is our propensity to attribute our inherent attitudes, stereotypes and expectations based on experiences and knowledge, which is.
a fundamental bias, should be taken into account when gathering eyewitness testimonies in real-world cases (Hellmann & Memon, 2016; Laney & Loftus, 2013). Additional notion in reported inaccurate details (in free and cued recall) to shed light on is the proportion of participants unable to accurately describe the appearance and clothing of the man and the woman in the filmed event. An important notice is all participants’ cognitively beneficial circumstances when witnessing the filmed event in contrast to real-world cases. Real-world crime cases entail inhibiting estimator variables such as circumstances surrounding the exposure (e.g., lighting, angle, distance, duration of exposure), presence of distractions and disturbances (such as weapons and noises) and the internal state of the witness (e.g., attention, motivation, skills and prejudices; Albright, 2017). However, despite beneficial circumstances during encoding phase, a substantial proportion of participants still managed to remember inaccurately regarding the appearance and clothing of both the man and the woman. Illustrations of reported inaccurate descriptions were the man with long hair when the man is clearly bald, another is that the woman is blonde when clearly, she is red-haired. Corroborating earlier findings, such memory errors and inaccurate detail information can in real-world crime cases result in ramifications, such as wrong offender signalments, which in turn can lead to inhibited and prolonged investigation. In a more drastic scenario, innocents can also be affected and in worst case scenario deprived of freedom (Albright, 2017; Shermer, Rose, & Hoffman, 2011).

In condition Second language through interpreter, the free recall was carried out with the witness providing testimony through an interpreter with a consecutive interpretation approach with direct style of interpretation (first person rendering) recommended by Maddux, (2010). Due to lack of participants applying for the role of interpreter, same interpreters were reused. Reusing interpreters enables potential inference and transfer of accurate and/or inaccurate detail information or conceptions between witnesses (e.g., source monitoring or misinformation effect; Johnson, Hashtroudi & Lindsay, 1993; Loftus, 1979), henceforth, potentially detriment the testimonies. However, this was to some sense controlled for with the cautious and precise informative description of the task in demand, especially with imposed verbatim rendering. A further notion in condition Second language through interpreter was the active language adaptation of witnesses in favor of the interpreter. The spoken language was basic and descriptive which facilitated the interpreters task. However, adaptation of language could possibly inhibit search paths in memory, consequently, hindering the prosperity of detail information and expanded perspectives. Moreover, research shows that actively constructing adaptation of descriptions non-fluently could exhaust the cognitive resources (ego-depletion) required for accurate recall of a crime event. Increasing the reliance on stereotype-consistent information when recollecting memories, specifically of distressing event, which consequently is potentially detrimental for an ongoing investigation (Harkness et al., 2015; Osborne & Davies, 2013). A further notion enclosing the conditions Second language and Second language through interpreter, research (Service, Simola, Metsänheimo, & Maury, 2002; Smith, Multhaup and Ihejirika, 2017) supports exhausting of cognitive resources when switching language, subsequently supporting the possibility that a state of ego depletion may affect resources utilized in recalling a crime event in a second language, consequently, affecting the accuracy and the extent of provided detail information in eyewitness testimonies. Still, the results in the present study showed no significant difference between all three conditions in the extent of provided detail information (accurate or inaccurate). Hence, the overall findings of the present study fail to support previous findings on ego depletion vulnerability to contaminations (Otgaar, Alberts and Cuppens, 2012a; 2012b; Szpitalak and Polczyk, 2014).
Limitations and future directions

An initial and substantial limitation with the present study was that the one-week retention interval had to be excluded after testing 16 witnesses due to complications (such as time pressure, missing data and lack of application). In the present study, twelve participants in condition Native language, twelve participants in condition Second language and four participants in condition Second language through interpreter had a retention interval of 6 to 8 days. The remaining participants across all three conditions completed the total experiment at one occasion, hence they were not exposed to potential contamination of memory between encoding and retrieval phase (Ecker, Lewandowsky, Cheung, & Maybery, 2015; Shepherd, 1983). This notion could possibly explain the ample extent of accurate detail information and the small proportion of reported inaccurate detail information across all memory tests in all three conditions. In line with the excluded retention interval, no filler task was substituting the effect of a retention interval, consequently the procedure did not adequately correspond to a real-world crime event. However, when observing mean values (and standard deviations), a large difference was evident between participants with a one-week retention interval in contrast to participants without the retention interval. Verifying the limitations with the present study and corresponding to real-world cases that imply retention interval between encoding of a crime event and retrieval when providing eyewitness testimonies during interrogation (Kaplan et al., 2016).

An additional important aspect to mention is the total amount of participants in the present study. Through a power analysis perspective, a total of sixty participants across three conditions (n = 20) is a small dataset to derive reliable statistical analyses from. The results presented in the present study should be perceived as tendencies towards the investigated variables. Therefore, there is a fundamental error which should be taken into account.

Moreover, the language variable was obstructed across three conditions (Native language; Second language; Second language through interpreter) to simulate language barriers on eyewitness testimonies to some extent. However, participants assessed in second language were all students at the Department of psychology, Stockholm university. Imposing a high level of language capacity for their second language English, consequently impeding the validity of the language barriers in the present study. Therefore, to improve the validity in the present design, future studies on language barriers on eyewitness testimonies should assess non-student participants or to a greater extent control for language capacity for the assessed second language by measuring language capacity with language tests.

Furthermore, from a participants’ perspective, during the encoding phase of the mock crime event when participating in an expected memory experiment, participants’ environmental circumstances was favorable in contrast to real-world cases. Suggesting a decrease in potential memory contaminations and distortions. Additionally, cognitive capacity was to some extent controlled for with participants self-reported assessments. The results from the self-reported assessment on memory ability, ability to convey memory, contextual valence, motivation and nervousness, showed no significant difference between conditions and no effect on the provided eyewitness testimonies. This is in substantial contrast to real-world crime events regarding the above-mentioned estimator variables (Albright, 2017). Research shows, when witnessing emotional intense events (such as crime events), eyewitnesses undergo a cognitive narrowing of the scope on acquired detail information towards features and details relevant to current intrinsic goals. The cognitive reduction of attention, when engaged in emotional intense events, can potentially increase the vulnerability and susceptibility of memory distortions, consequently causing overall memory deficits specifically on non-goal relevant details.
Furthermore, Kaplan et al., (2016) posits that eyewitnesses to crime events often is exposed to a retention interval between the crime event and the interrogation, consequently, increasing the possibilities of potential contaminations on the narrowed and malleable memory. Consequently, confirming the notion of methodological errors in the present study, which subsequently offers explanations to the acquired non-significant results in the present study.

Furthermore, potential limitations concerning the stimulus material could explain the extent of accurate eyewitness testimonies. A further investigation ought to examine self-report on experienced distress by the specific mock crime event used in the present study. A library setting encloses an inherent non-distressing mindset which feasibly could prime cognitive abilities which subsequently facilitates the encoding phase. A comparison between a non-distressing mock crime event and a distressing mock crime event ought to be examined and whether it affects the accuracy of the eyewitness testimonies. Furthermore, despite reported high motivation in the present study, there could be a substantial difference between high motivation in laboratory studies on crime events and high motivation in real-world crime events, potentially affecting the consolidation of detail information in memory and willingness to contribute to closing criminal investigation. Therefore, future studies ought to investigate the differences on motivation level and eyewitness testimonies between laboratory studies and real-world events to avoid draw erroneous inferences between laboratory studies and real-world events on crime events.

Additional limitation with the present study is that the participants were psychology students. Inherent insights on psychological phenomenon tend to reduce subsequent biases. Furthermore, the experimenter effect in behavioral science should be illuminated. Psychology students are indoctrinated in the constructs of psychological experiments and henceforth in a greater extent attentive towards identifying task demands, developing beliefs about the researchers’ hypotheses and consequently adopt more. Therefore, suggesting a contrast between psychology students and non-psychology students with the present design as future investigation.

An interesting follow-up study on the present study and to further study the credibility-accuracy discrepancy and biases, would be to investigate judgments on credibility of eyewitness testimonies as done in Lindholm (2008). Given that the present study obtained results suggesting that interpreters judged credibility higher when witnesses reported a greater proportion of detail information than when witnesses reported a low proportion of detail information, regardless of the extent of provided accurate and inaccurate detail information. Positing, in line with Lindholm (2008), low validity in judgment of credibility, which could cause ramifications in the legal system due to judgements of high witness credibility. Furthermore, an additional investigation would be to ask participants to judge the man’s guilt in the mock crime event. Then, participants would re-evaluate their recollection of the fragmental crime event, with subsequent potential contaminations based on stereotypes, expectations, attitudes and moral. Hence, re-testing the role of social desirability and interrogative suggestibility as described by Richardson & Kelly (2004) and Gudjonsson (2007).

Further intriguing directions for the future regarding applied witness psychology is to examine individual differences, to further deepen and broaden the knowledge for investigators to apply when investigating eyewitness testimonies (e.g., motivational aspects behind providing an eyewitness testimony, differences in personality traits moderating the eyewitness testimony, and whether previous experiences potentially moderate the capability of providing eyewitness testimony positively or negatively). Evidently, extensive research on applied forensic psychological research on eyewitness psychology can be conducted.
Conclusion
In summary, the present study yielded inconclusive results with no effects of language barriers on accuracy and suggestibility in eyewitness testimonies. However, results from the two language barriers conditions, Second language and Second language through interpreter, tended to go in the same direction, suggesting trends of effects of language barriers on accuracy and suggestibility. Such tendency and increasing linguistic diversities in our society, cultures and sub-cultures are yielding a potential risk towards systematic discrimination in our legal system with likelihood of encountering inaccurate and confabulated testimonies when faced with a critical forensic interrogation process. Therefore, further investigation on the topic with rigor and robust method should be conducted to test whether the trends towards negative effects of language barriers are evident and worrisome. Clearly, more applied forensic psychological research on eyewitness testimonies is needed.

References


https://doi.org/10.1177/1043986211405886