A little birdie told me:

People will consider ambiguous utterances from parrots when making judgements of guilt.

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Science (Research) in Psychology

at

The University of Waikato

by

Callum Daniel Lawrie

2020

Abstract

How much would you trust a parrot as an eyewitness? The question is not hypothetical: lawyers have tried to use "testimony" from animals. We[1] asked two questions: First, had a court allowed a parrot's utterance to be entered as evidence how credible would people believe the parrot is? And second, when parrots give evidence that is ambiguous, what does that evidence do to people's judgements of guilt? To answer these questions we ran three experiments. Subjects were presented with a vignette describing a husband returning home to find his wife murdered and were informed that either a parrot or a three-year-old child was repeating an ambiguous utterance following the murder. We found that people rate a parrot as less credible than a three-year-old child but do not dismiss the parrot entirely. We also found that people can be persuaded by a parrot's ambiguous utterance when making a judgement of guilt. These results suggest that people will consider ambiguous utterances made by less than credible sources when making judgements of guilt.

---

[1] Although the research in this thesis is my own, I conducted it in a lab and was supported by a team comprised of a postdoctoral research fellow, PhD students and honours students. I also received advice and direction from my supervisors. Therefore, I often use the word "we" in this thesis to reflect that fact. As you will also see, I use the word "we" in a different context to refer to what is known (or not known) in the wider scientific community.

Acknowledgements

~

"I have found, through painful experience, that the most important step a person can take is always the next one." — Brandon Sanderson

Table of Contents

List of Tables and Figures

A little birdie told me: People will consider ambiguous utterances from parrots when making judgements of guilt.

In 2015, a woman in Michigan, USA, stood accused of murdering her husband before attempting to take her own life. As the case unfolded, the police discovered an unusual witness. The husband's pet parrot had been repeating the utterance "Don't fucking shoot." A prosecutor in the case discussed the possibility of using the parrot's utterance as evidence in the trial (Siemaszko, 2016). The case in Michigan is not the only case in which a parrot has repeated an utterance that might have been related to a crime. In 1993, a murdered woman's pet parrot, who had been left in the care of a local pet store, repeatedly uttered "Richard, no, no, no!" The suspect in the case was named Gary Rasp, not Richard. Mr Rasp's defence lawyer attempted to introduce the information as evidence to support his argument that Mr Rasp did not murder the woman, but an objection from the prosecutor was sustained by the judge and the defence was unable to finish introducing the parrot's utterance. Ultimately, Mr Rasp was convicted of murder ("Parrot may have the answer to a killing," 1993). In 2014, a case in India was solved following a "lead" from a parrot. Local police had been unable to solve the murder of a wife and her pet dog—until the husband of the deceased woman noticed that their pet parrot would become fearful when their nephew visited. The husband became suspicious and repeated the names of potential suspects to the parrot. When the husband mentioned the nephew, the parrot would become agitated and squawk "Usne mara, usne mara (he killed, he killed)." Following this "lead," the police arrested the nephew and ultimately obtained a confession (India Today Online, 2014).

These examples led us to ask two questions: First, had a court allowed a parrot's utterance to be entered as evidence, how credible would people have believed the parrot was? And second, if parrots give evidence that is ambiguous—much like a parrot saying "Richard, no, no, no!" When the defendant's name is Gary, what does that evidence do to people's judgements of guilt vs. innocence? When we look to the psychological literature for answers to these questions, we see that there is reason to expect that people would recognise that parrots are less-than-credible witnesses and would not consider evidence from a parrot when deciding someone's guilt. But there is also reason to expect that

people may rate a parrot as credible and would consider evidence from a parrot when deciding someone's guilt. To address these issues, we conducted three experiments to examine how credible people believe parrots are and whether people incorporate evidence from a parrot when judging someone's guilt.

Parrots and other non-credible witnesses present a number of problems in the courtroom. First, information obtained from a parrot or similar witness could be interpreted any number of different ways because the information is considered out of context, and we cannot obtain any new information. It is not possible, for example, to cross-examine a parrot. Therefore, if a parrot repeated the utterance "John don't," we cannot ascertain when the parrot overheard the phrase, whether it had heard the utterance in different situations, or who it heard speak the phrase. Second, the information is out of context and therefore ambiguous. Did the utterance "John don't" refer to John not shooting his wife or John not tickling his wife? Given that we do not know what John should not do, jurors in a courtroom can only guess at the context and the meaning of the ambiguous utterance "John don't."

If people could dismiss ambiguous information, it might be less concerning. But ambiguous information can affect the outcome of a trial. For instance, we know people interpret ambiguous statements depending on their own personal experience and their current state of mind (Dearing & Gotlib, 2009; Lau et al., 2011). People will, therefore, interpret ambiguous statements to fit a story that makes sense based on their prior experience. For example, a common trope in crime shows is that the husband is usually the killer. The trope probably exists for good reason: over one-third of women murdered are killed by their husbands or boyfriends; men are also six times more likely to murder their intimate partner than women (Hanlon et al., 2016; Stöckl et al., 2013). It seems reasonable, then, to think that people often believe the husband is the killer. Because the utterance "[husband] did it" is consistent with this familiar theme, people may be inclined to interpret that utterance as relevant to the crime. Parrots further exacerbate this problem because we cannot ask for additional details to resolve the ambiguity. Ambiguous information also presents a problem in a court room setting. People can be encouraged to interpret ambiguous information in a particular way depending on individual factors such as the context in which people encounter that information and their past experiences

(Duffy & Feist, 2014). Therefore, we might expect that because the ambiguous information is presented in a courtroom people would be encouraged to interpret that information as evidence and would not consider the fact that the statement is ambiguous.

Another problem is that people also fail to recognise that a statement is ambiguous, particularly after having drawn a conclusion about what that statement means (Wylie et al., 2019). For example, a juror may use evidence that was deemed ambiguous by the judge and, consequently, inadmissible. But because jurors are encouraged to interpret information presented in a courtroom as evidence, they will still use that information when deciding their verdicts (Duffy & Feist, 2014). If a parrot uttered something ambiguous, people might not even notice the ambiguity. As a result, they would still interpret that utterance to fit a familiar theme as described earlier.

But before we investigate whether people would be persuaded by a parrot's ambiguous utterance, we first need to consider the extent to which people would even consider a parrot's utterance when making a judgement of guilt.

Reasons to believe people would not be persuaded by a parrot when making a judgement of guilt

In an ideal world, people would never consider information they should not use when making judgements of guilt and, therefore, would not use information provided by a parrot. A typical pet parrot is not a credible source of information. Although we have some evidence that suggests that an African grey parrot could produce novel speech with enough training this ability would not be common for the average pet parrot. Alex the parrot was able to identify individual phonological units and recombine these units to produce novel speech (Pepperberg, 1999; 2007). This level of training is not typical for a pet, Alex spent 26 years receiving intensive interspecies communication training, even younger parrots undergoing the same training are unable to produce novel speech (Pepperberg, 2007). Therefore, it seems more plausible that a typical pet parrot would, at most, only be able to mimic overheard phrases. Furthermore, because we are unable to establish the context of these phrases, any information provided would be highly ambiguous. Intuitively, the idea that jurors would be persuaded by a pet parrot's utterance may seem absurd. We might expect that, because a parrot is unlikely to be

considered a credible witness, people would dismiss any information provided by a parrot. After all, research suggests that the credibility of a witness is related to the persuasiveness of that witness (Brodsky et al., 2010; Goodman-Delahunty et al., 2010; Voogt & Klettke, 2017; Voogt et al., 2020). Furthermore, we know that people will dismiss witnesses if it becomes clear that they are more confident than they should be (Tenney et al., 2011; Tenney et al., 2008). For example, a witness claiming to be confident that he saw the suspect commit the crime over 100 meters away, at night, in poor lighting, would probably be seen as dubious.

People can correctly dismiss misinformation. In the misinformation literature, we see people can correctly dismiss false information from non-credible sources. People will actively discount misinformation from sources that are likely to be biased, such as a defence lawyer (Dodd & Bradshaw, 1980). Likewise, people will assess their own relative credibility compared to others before accepting or rejecting misinformation (French et al., 2011). Moreover, people can dismiss misinformation after being introduced to it once they are informed that the source may not be credible (Echterhoff et al., 2005). Therefore, if people recognise that a parrot is not a credible witness, they should be able to ignore the evidence presented by a parrot and will not be persuaded when making a judgement of guilt. Even children will consider credibility before using information from a source. Children will assess an adult's past reliability and accuracy before accepting new information from that adult, rather than merely accepting that information as accurate (Jaswal & Neely, 2006). Taken together, this body of work suggests that people are quite good at distinguishing unreliable information and subtracting it out of their evaluations. As far as parrot witnesses go, if people are able to recognise the parrot is less than credible and doubt its accuracy, then we would expect that they would not be persuaded by the parrot's evidence.

Reasons to believe people would be persuaded by a parrot when making judgements of guilt

Despite all the reasons above, even if a parrot were considered non-credible, there are also a number of reasons to think the parrot's utterances might influence people's judgements of guilt vs. innocence. For example, we know that although people are able to dismiss evidence when they do not

believe the source is credible, they still mistakenly are persuaded by information or evidence they should not be during decision-making. This concern is not hypothetical. Jurors are often exposed to evidence which should not be used when forming judgements of guilt (Kassin & Sommers, 1997). This evidence includes evidence considered inadmissible, pretrial publicity, and the "story," rather than the actual facts, of the case.

People consider inadmissible evidence. Inadmissible evidence is evidence which is not to be considered when making judgements about a case. For example, inadmissible evidence can include hearsay, evidence obtained incorrectly, and inappropriate witness testimony. Judges cannot control every statement made in the courtroom to ensure that no inadmissible evidence is introduced. Therefore, courts expect people to ignore any evidence that is deemed inadmissible after it is presented (Daftary-Kapur et al., 2010). But even if jurors are explicitly told to ignore inadmissible evidence, they often will still consider that information when making decisions. The ability to actively track inadmissible evidence and then to intentionally exclude that information when forming a judgement requires a considerable amount of cognitive effort (Daftary-Kapur et al., 2010). A meta-analysis conducted by Steblay et al. (2006) reviewing 175 experiments found that inadmissible evidence still influences verdicts. Moreover, people will still consider the inadmissible evidence even when given a specific judicial instruction to ignore that evidence in their judgements (Demaine, 2008; Mazzella & Feingold, 1994; Steblay et al., 2006). One reason that people cannot ignore inadmissible evidence is likely due to the paradoxical effect of thought suppression. When tasked with ignoring a piece of information, instead of not thinking about it, people instead will think about that information more (Abramowitz et al., 2001; Wegner et al., 1987). Therefore, even if people acknowledge that they should not be persuaded by evidence from a parrot, or are even explicitly told not to consider that information as evidence, they might still ultimately be persuaded by a parrot.

People still consider pretrial publicity. Inadmissible evidence is not the only type of information people consider when they should not. Potential jurors are often exposed to information outside of the courtroom in the form of negative pre-trial publicity via the media, which portrays the suspect as guilty. When people become members of a jury, they are expected to discount any

information they have been exposed to before the trial. But jurors are frequently unable to disentangle trial evidence from pre-trial publicity. Jurors who have been exposed to negative pre-trial publicity are more likely to vote a suspect guilty, have a more negative impression of the suspect, and recommend harsher sentences (Daftary-Kapur et al., 2014; Ruva et al., 2007; Steblay et al., 1999). Furthermore, the effect of pre-trial publicity lasts for an extended period, such as the time between reading about a case and being called for jury service (Daftary-Kapur et al., 2014). Not only do jurors show bias when exposed to negative pre-trial publicity, but they often cannot distinguish when they are being influenced by it. What is more, jurors often make source monitoring errors, showing an inability to distinguish where or from whom they received information, and struggle to identify whether the information has come from media articles or the trial itself (Ruva et al., 2007; Ruva & Guenther, 2015; Johnson et al., 1993). Even more concerningly, people who have not been exposed to pre-trial publicity can still be influenced during deliberation by others who have been exposed (Ruva & Guenther, 2017). People often struggle to distinguish between different sources of information in a courtroom and fail to judge what information should and should not be used when making judgements. Again, we see evidence that people will consider information from a variety of sources when making a judgement of guilt regardless of whether they should or should not.

Taken together, it is well established that people often use information when making judgments that they should not — such as inadmissible evidence, and information obtained outside of the courtroom. Therefore, it is not implausible that people may, in fact, be persuaded by an utterance provided by a parrot.

People construct a story to understand a case. The most commonly accepted model of jury decision making is the Story Model (Pennington & Hastie, 1992; Willmott et al., 2018). The Story Model proposes that jurors will consider and evaluate evidence by constructing one or more plausible stories of what happened. Jurors will then consider the possible verdict categories and will use prior knowledge to categorise potential verdicts. Finally, jurors will assess which story fits best and will match that story to the verdict that fits best (Pennington & Hastie, 1991, 1992). People are more likely to rate a suspect guilty when the suspect makes sense in a juror's story. Additionally, people are more likely to

have a negative view of a suspect and rate the suspect as guilty when the suspect fits a stereotype of the sort of person who would commit that crime (Jones & Kaplan, 2003). Similarly, people are more critical of a case if a suspect does not fit a stereotype, and they will be more likely to question the validity of the claim (Schuller et al., 2010). For example, a black male charged with an assault would be more likely to be found guilty, compared to a young woman accused of sexual assault. Not only do people judge a suspect based on how stereotypical the crime is, but the victim's response style can also influence a juror's decision. Women who show the stereotypical emotional responses that are expected when describing a crime are more likely to be believed than women who do not (Hackett et al., 2008; Kaufmann et al., 2003; Nitschke et al., 2019; Schuller et al., 2010). For example, a crying woman describing an assault who is visibly shaken and upset is more likely to be believed than a women who is cold and calm describing the same assault. It is, therefore, plausible that if a non-credible witness provided evidence that fitted with a juror's story, the juror may be persuaded by this evidence when making judgements.

    People are less critical of information that fits their story. Not only are people more likely to be persuaded when information fits with their beliefs, they will also change how critically they analyse information depending on whether or not it fits with their beliefs (DeKay, 2015; Russo, 2014). In other words, instead of fairly analysing each new piece of information to construct their story, once people have started to form a theory they will interpret the evidence as either in support of or against their theory. This phenomenon is known as *predecisional bias*. We know that once people have made a particular decision that they will be more critical of information that goes against their decision and less critical of information that supports their decision (DeKay, 2015; Russo, 2014). We see a similar effect in the confirmation bias literature: People will be more critical of information that does not support their belief compared to information that does support it (Nickerson, 1998; Taber & Lodge, 2006). People are even more prone to confirmation bias when information is presented sequentially, compared to when all information is presented simultaneously. Of course, information presented in a courtroom is always sequential, which further increases the likelihood that people will be less critical of information that fits with their belief (Jonas et al., 2001). It is, therefore, plausible that people may

be less critical of a parrot's utterance if that utterance fits with their story of the crime. For example, if people already think that the husband murdered his wife and are then informed that the parrot is repeating the utterance "John don't," they would be less critical of that information than if the utterance was "Sally don't."

People assess credibility using non-credibility-related cues

The credibility literature suggests that people judge credibility across five factors: How accurate is the witness likely to be, how believable is the witness, how reliable is the witness likely to be, how competent is the witness, and how truthful does the witness appear to be (Voogt et al., 2019; Voogt Klettke, & Thomson, 2017)? But we know that people will consider cues which do not accurately indicate overall credibility to judge credibility. For example, people often will assess confidence as a shortcut to assess someone's credibility; this shortcut, called the *confidence heuristic,* operates so the more confident someone appears, the more credible people believe she must be (Pica et al., 2019; Price & Stone, 2004; Sah et al., 2013). Using confidence may make sense intuitively; the witness must be confident because she saw the event and can accurately recall the details.

But confidence has little to do with overall accuracy or actual credibility. People can have a completely false memory and still show confidence in their memory, describe their memory with fine details, and express strong emotions despite the memory being completely false (Loftus, 2003). Even when people recall memories that are not false, people's confidence in their own memory is highly malleable (Luus & Wells, 1994). Because people assess confidence as a shortcut for assessing credibility, confident speakers will hold more sway in a courtroom regardless of their overall accuracy (Pica et al., 2019; Sah et al., 2013). People do not only consider how confident someone claims to be to assess credibility, but people will also consider the way in which a person speaks when assessing credibility. If a witness employs more hedging words, such as,"I think" or "I believe" people will rate the witness as less credible and will doubt their knowledge, regardless of his overall accuracy (McCarthy Wilcox & NicDaeid, 2018; Ruva & Bryant, 2004). In addition to how confident a witness appears to be, people will also consider a range of other non-credible cues such as how kind, friendly, pleasant, respectful, and well-mannered a witness is to judge credibility (Brodsky et al., 2010). Therefore, even when people

scrutinise a witness's credibility before assessing his statement, they often consider non-credible related cues including, but not limited to, confidence, likeability, and speech style to judge a witness's credibility.

Moreover, people might find pets credible because they form strong attachments to their pets. A 2010 poll conducted by Harris Decima polling found that 53% of people believed that their pets were more reliable than people (Harris/Decima Inc, 2010). Furthermore, 91% of pet owners consider their pets as members of their families (Harris Interactive, 2012). It is plausible that people would draw on their own attachment to their pets when assessing credibility, even though their attachment is not very relevant. After all, we know people often draw on irrelevant feelings during a situation to make decisions in that situation (Schwarz, 2012). Therefore, we might find that people rate a pet parrot as somewhat credible.

People are not always persuaded by the more credible source

Not only is it plausible that people will rate a parrot as somewhat credible, it is also plausible that people could be persuaded by an objectively less credible witness if they believe that witness would be more truthful. When we look to the child witness literature we find evidence that people are persuaded by less credible witnesses compared to more credible witnesses. Children are often asked to provide eyewitness accounts, particularly in situations of sexual and physical abuse, in which the only evidence against a suspect can be the child's testimony. Although a child can be accurate and credible under the right circumstances, how accurate a child is when recalling information depends on how the court obtains the information from the child, and the child's cognitive ability—how advanced the child is for their age (Zajac & Hayne, 2003; Zajac et al., 2012). Children are typically cross-examined in court in a manner that has been shown to decrease accuracy (Zajac et al., 2012). Children are also particularly susceptible to suggestion from parents and authority figures (Ceci & Bruck, 2006; Pantell, 2017). Children are also more likely to answer ambiguously-worded questions than adults (Zajac et al., 2003). Due to the difficulty of obtaining accurate information from children, and their increased susceptibility to suggestion, evidence from children should receive more scrutiny as their credibility cannot be assumed based solely on how credible they appear to be.

Despite the need for caution, people generally trust children. People are more likely to rate a young child as more credible than a young adult (Bottoms et al., 2004; Golding et al., 2015; Rogers et al., 2007; Ross et al., 1990). In fact, people tend to rate younger children as more accurate, believable, reliable, and trustworthy than older children (Voogt et al., 2020). People appear to believe that younger children are more likely to be honest and therefore are more believable (Ross et al., 1990). But younger children are not inherently more truthful than older children. Children as young as four years old are willing to lie to protect an adult or themselves (Bottoms et al., 2002; Lyon et al., 2008). Moreover, people are more willing to consider information from a younger child and will rate the younger child as more credible than an older child, even though an older child is objectively more likely to be credible (Zajac & Hayne, 2006).

Taken as a whole, we have evidence that people can be persuaded by how truthful they believe a witness is regardless of the witness's overall credibility. We also know that people believe their pets are reliable (Harris/Decima Inc, 2010). Pet parrots are unlikely to lie given that they can only repeat information and cannot produce novel speech (Pepperberg, 2007). It is therefore, plausible that people would also believe that parrots are truthful and would be persuaded by evidence given by a parrot.

To summarise, the literature suggests a number of factors might come in to play if people were asked to consider the credibility of a parrot's utterances. For example, we know people often fail to recognise ambiguity. We also know that jurors often consider information they should not, such as inadmissible evidence, and fall sway to the influences of pretrial publicity and biases. Moreover, people can be persuaded by less-than-credible witnesses. But what we do not know is to what extent people might consider a parrot a credible witness—and if so, how credible, especially when compared with a young child whose mental capacity could be considered similar to that of a parrot (Singh et al., 2017; Schloegl et al., 2012; Pepperberg, 1999).

To address these issues, we asked subjects to consider the murder of a woman found dead in her home. Her husband claims to have come home and discovered her body. Some subjects read that the husband's pet parrot is repeating an ambiguous phrase that could, if taken one way, implicate the husband. Other subjects read that the couple's young child is repeating the same phrase. How do

subjects then rate various measures of credibility with respect to the "witness," and the guilt of the husband? Because people consider non-credible cues when assessing credibility and believe pets are reliable we expected that people would rate a parrot as somewhat credible. Additionally, people will consider evidence that should not be considered such as inadmissible evidence. Moreover, once people construct a belief of what happened they are less critical of information that fits with that belief; we therefore expect that people will be persuaded by ambiguous utterances from less-than-credible sources when making a judgement of guilt.

We addressed these issues in three experiments. Experiments 1 and 2 established how credible people believe a parrot is compared to a three-year-old child. Experiment 2 established relative ratings of guilt. Experiment 3 then manipulated the ambiguity of the utterance provided by the parrot. Across these experiments, we found evidence that people believe a parrot is less credible than a three-year-old child, but do not dismiss the parrot's utterance entirely. We also found evidence that people will consider ambiguous utterances from the less-than-credible parrot when making judgements of guilt.

Experiment 1

The primary purpose of this experiment was to determine the extent that people believe a parrot is a credible witness, relative to a three-year-old child. We selected a child witness as parrots have a similar reasoning ability to a three-year-old (Singh et al., 2017; Schloegl et al., 2012; Pepperberg, 1999). Recall that people judge credibility across accuracy, believability, competence, reliability and truthfulness (Voogt et al., 2019; Voogt, Klettke, & Thomson, 2017). Because people believe that pets are reliable and often fail to recognise ambiguous statements, we predict that people will rate the parrot as somewhat credible, although less credible than a three-year-old child.

Method

Subjects. Subjects were recruited via Amazon's Mechanical Turk (MTurk), an online community of workers who volunteer to carry out various tasks. MTurk is commonly used by researchers to collect data (Mason & Suri, 2012). A total of 204 American Mechanical Turk workers completed the experiment in exchange for 0.30 USD. The subjects' age ranged from 18 to 83 years,

(M = 42.05, 95% CI [40.14, 43.96], SD = 13.83); 124 subjects identified as women, 79 identified as men, 1 identified as non-binary; 94.11% of subjects reported English as their first language, and 100% reported English as their primary language, the language they use most day-to-day.

Design. We used a between-subjects design with two conditions (witness: three-year-old child, African grey parrot).

Procedure. The subjects completed the experiment online via Qualtrics, an online survey program (https://www.qualtrics.com). Subjects provided consent before beginning the study. We told subjects that the study was examining learning styles to ensure that subjects were unaware of the true purpose of the study. This minor deception was necessary in order to prevent participants from guessing the true purpose of the experiment and altering their responses as a result. We asked that subjects completed the experiment in lab-like conditions. For example, subjects were asked to complete the experiment in one session, somewhere quiet, and without the assistance of anyone else.

Subjects read a vignette describing a murder with no suspects or evidence linking anyone to the crime.

> Two weeks ago, Mr. [Husband's name] Wilson contacted the police, saying that he had come home from work at 6 pm to find his wife in the backyard, covered in blood and dead. An autopsy showed that Mrs. Wilson had been shot in the chest, and had quickly bled to death sometime between 1 pm and 3 pm. The police found no evidence at the scene to link the murder to any specific person.

Subjects were then told that the victim's son / parrot (Leroy) was repeating the utterance "[Husband's name] did it" and that investigators were considering using the utterance as evidence. Subjects were told that the husband's name was either Paul or Charlie. We selected two equally-common first names for middle-aged men in the United States of America (Social Security Administration, 2019).

Subjects then completed an adapted Child Sexual Assault Victim Credibility Scale (CSAVCS) as a measure of credibility. The CSAVCS is a scale designed to measure the perceived credibility of child sexual assault victims when testifying in court. The scale measures five subscales: accuracy, believability, competency, reliability, and truthfulness across 23 items which when combined provides

an overall credibility score. We chose this scale because it has good psychometric properties; each subscale has good internal consistency ($\alpha_{accuracy}$ = .84, $\alpha_{believability}$ = .86, $\alpha_{competency}$ = .78, $\alpha_{reliability}$ = .82, and $\alpha_{truthfulness}$ = .92; see Voogt, Klettke, & Thomson, 2017 for a review). We adapted the CSAVCS to fit both a parrot and three-year-old child (see the Appendix B; Voogt, Klettke, & Thomson, 2017). For example, "This child's report was believable" was modified to "Leroy's report was believable" to include both witness conditions. Additionally, we removed two items which did not apply to our vignette. Because we adapted the CSAVCS, in addition to answering the research question, we also gathered evidence to check that the scale was measuring the same construct by running a principle component analysis; these data appear in the results.

On the CSAVCS, the mean credibility score is obtained across all 21 items to produce an overall score (Voogt, Klettke, & Thomson, 2017). Each item was presented in random order to the subjects. For example, subject one may answer each item in one order (1,2,3, etc.) while subject two may answer each item in a different order (2,1,3, etc.). For each item, subjects used a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) to report their agreement with the credibility statements. In an open text box, subjects then described how they believed the crime unfolded. Finally, subjects were asked to complete a series of yes / no questions about their compliance with the earlier request to complete the study in lab-like conditions.

Results and Discussion

Recall that the primary purpose of this experiment was to determine the extent people believe a parrot is a credible witness, relative to a three-year-old child. Before addressing our research question, we checked for subject responses that included nonsensical descriptions of the crime or did not use English as their primary language. We did not find any subject responses that met these criteria and therefore proceeded without any exclusions.

CSAVCS principal component analysis. We checked for evidence that the adapted items in the CSAVCS measured the same construct reliably. We ran a principal component analysis of each survey item to check that each item was still measuring a single construct, credibility. We found that all items except for four indeed loaded on this single factor. These data appear in Table 1.

The other four items loaded on an additional factor. As shown in Table 1 these four items were: "[Witness] is not believable", "[Witness] is not a credible witness", "[Witness] is not a competent witness", and "[Witness]'s memory of the event is likely to be unreliable" of which only "[Witness] is not a competent witness" loaded on an additional factor more than the first factor. To assess whether "[Witness] is not a competent witness" should be removed, we examined the Cronbach's α, which was 0.963 (Cronbach, 1951). When we removed "[Witness] is not a competent witness," the overall alpha increased to 0.964 and so we retained the item as the scale's overall reliability would not have improved. Because our adapted CSAVCS was loading on a single construct and showed high internal consistency reliability we concluded that the adapted scale was measuring a single construct.

Credibility. We now return to the primary question. To address the issue of how much people believe a parrot is a credible witness, relative to a three-year-old child, we first calculated the mean credibility score of all 21 CSAVCS items, classified according to whether the witness was a parrot or a child. We then plotted those means in Figure 1.

As Figure 1 shows, subjects rated a three-year-old child as more credible than a parrot. In Null Hypothesis Significance Testing (NHST) terms, we ran an independent samples t-test to examine the effect of witness type (three-year-old child, African grey parrot) on subjects' mean credibility ratings. Subjects rated the three-year-old child as significantly more credible than an African grey parrot (M_{diff} = 0.94, 95% CI [0.65, 1.23]).

In short, Experiment 1 provides two important findings. First, subjects rated a parrot as less credible than a child, but did not dismiss the parrot's evidence entirely. Subjects slightly disagreed with the credibility statements overall, which suggests that in some situations people may be persuaded by a parrot, although with some reservations. This finding fits our expectations.

Table 1

CSAVCS Adapted Items Principal Component Coefficient and Cronbach's a if Excluded

Item	Component			Cronbach's α if excluded
	1	2	3	
Leroy's utterance is accurate	0.848			.964
Leroy is able to give an accurate description of what happened to police	0.766			.964
Leroy's memory for the event is accurate	0.824			.963
Leroy's utterance is believable	0.837			.962
Leroy is not believable	0.665		0.541	.962
Leroy's utterance is credible	0.865			.962
Leroy is not a credible witness	0.556	0.485		.962
Leroy has the capacity to provide a competent account of events	0.777			.961
Leroy is not a competent witness *	0.499	0.566		.961
Leroy has a good memory for the event	0.762			.961
Leroy is reliable	0.791			.961
Leroy is dependable	0.836			.961
Leroy's account of events can be depended upon	0.818			.961
Leroy's memory of the event is likely to be unreliable	0.593		0.469	.960
Leroy would tell the truth about this event if he had the language ability	0.739			.960
Leroy's disclosure of what happened is truthful	0.849			.960
Leroy is honest	0.712			.960
Leroy's utterance is provided honestly	0.696			.960
Leroy is trustworthy	0.866			.960
I believe the event occurred as Leroy said	0.871			.960
I would trust a typical [child of this age / parrot of this species] to be telling the truth	0.851			.960
			α total	.963

Note. Coefficients only reported if greater than 0.4

* Item loaded on additional factor more than the first factor

Figure 1

Mean Agreement with 21 Credibility Items across Witness Conditions (parrot, child)
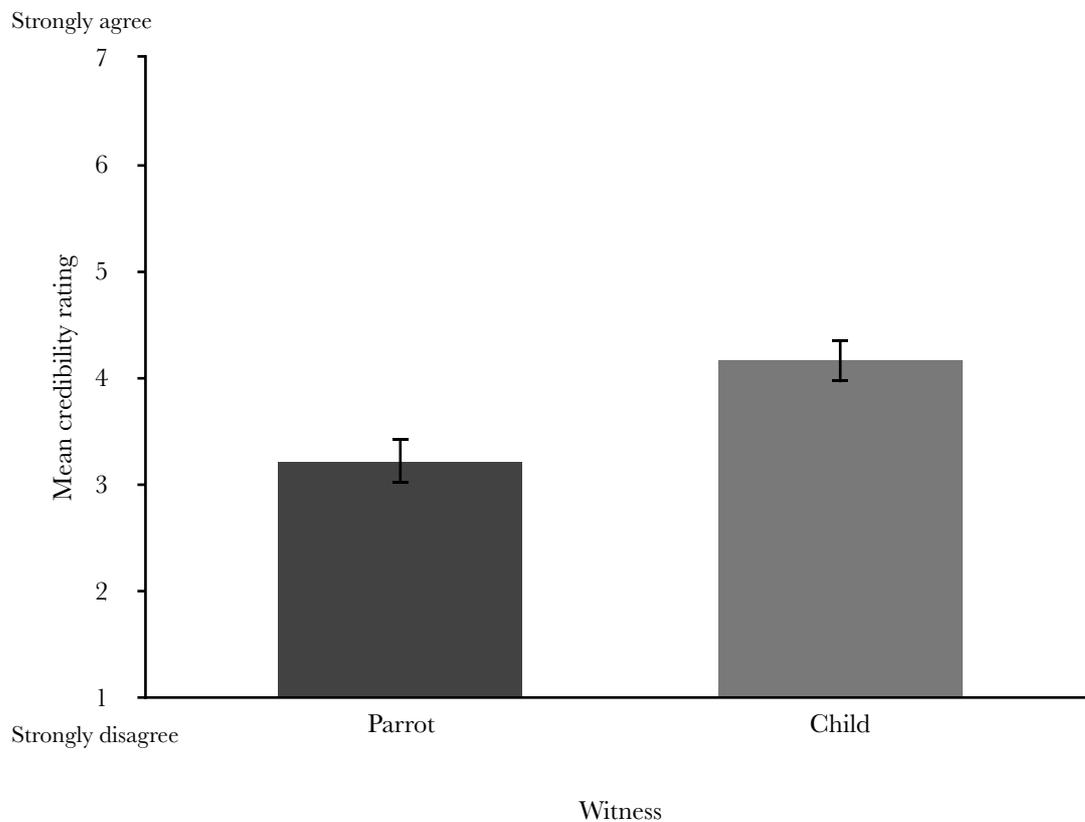
Note. Error bars represent 95% confidence intervals of cell means.

But there are at least three alternative explanations as to why subjects might have rated the parrot as somewhat credible. First, the CSAVCS anchors of *strongly agree* and *strongly disagree* was actually measuring agreement with the credibility statements, not credibility itself. That is, instead of telling us how credible subjects believed the witness was, the scale was telling us how much subjects agree with each credibility statement. Asking subjects how much they agree with the credibility statements may have led subjects to respond to each item differently than they otherwise might have had the scale measured credibility. Additionally, interpreting mean agreement with credibility statements is less clear than mean credibility. We address these problems in Experiment 2. Second, the truthfulness subscale in the CSAVCS might be inflating credibility ratings of parrots. Parrots are unlikely to lie as that would require an advanced level of cognition. Children only begin to lie around the age of three and four (Bottoms et al., 2002; Lyon et al., 2008). Although parrots have a similar

reasoning ability as three-year-old children, it is unlikely a typical pet parrot would be able to produce novel thoughts or to consciously lie (Mikolasch et al., 2011; Pepperberg, 2006). Therefore, subjects may be rating the parrot as somewhat credible because they believe the parrot is incapable or unlikely to lie and therefore is truthful. We address this possibility in Experiment 2 by analysing each individual subscale. Finally, subjects may not be aware of what a parrot is capable of (or not capable of) and believe that the parrot is producing unique thoughts, rather than repeating something overheard. We also address this possibility in Experiment 2 by adding four items to investigate how people are interpreting the utterance.

Experiment 1 cannot tell us anything about how people might actually consider information uttered by a parrot when assessing someone's guilt. We address this question in Experiment 2 by adding new items that ask subjects whether they believe the witness is repeating something overheard or unique thoughts, and whether they believe the utterance is related or unrelated to the crime.

<div align="center">Experiment 2</div>

Experiment 2 addressed three questions. First, to what extent do people believe a parrot is a credible witness compared to a three-year-old child? Second, to what extent do people use an utterance made by a parrot when making judgements of guilt compared to a child? Finally, to what extent do people rate a parrot's utterance as a unique thought or repetition compared to a three-year-old child?

We used the same method as Experiment 1 with changes as follows. An additional control condition was added in which there was no witness, and subjects only read the vignette pertaining to the husband. The anchor points on the adapted CSAVCS were changed to 1 = not at all, 7 = completely. We also added an attention check question to the scale (Voogt, Klettke, & Thomson, 2017). Finally, and most importantly, we added five new measures as described below. We preregistered this experiment (see Appendix C for the pre-registration).

Method

Subjects. Subjects were recruited via Amazon's Mechanical Turk. A total of 324 American Mechanical Turk workers completed the experiment in exchange for 0.30 USD. The subjects' age

ranged from 19 to 77 years, (M = 41.71, 95% CI [40.30, 43.11], SD = 12.88); 196 subjects identified

as women, 127 identified as men, 1 identified as non-binary; 94.13% of subjects reported English as

their first language, and 98.46% reported English as their primary language, the language they use

most day-to-day.

Design. We used a simple three-condition design, manipulating the type of witness (three-year-

old child, African grey parrot, no witness) between subjects.

Procedure. Subjects first were asked to read the same vignette from Experiment 1. Subjects in

the no witness condition were then asked to complete the scales; otherwise, subjects in the parrot and

child condition were then informed that a parrot or child respectively was repeating the utterance

"[husband] did it." In addition to counterbalancing the husband's first name (Paul and Charlie) we

also counterbalanced the husband's last name (Wilson and Anderson) two equally as common last

names in the United States (United States Census Bureau, 2016).

Subjects were then asked to complete the CSAVCS and to report whether they believed the

husband shot his wife on a 6-point Likert scale (1= definitely not, 6= definitely) as a measure of the

husband's perceived guilt. The order in which subjects completed the CSAVCS and measure of

perceived guilt was counterbalanced.

Subjects were also asked to rate four statements to determine whether they believed the witness

was repeating something or communicating his own unique thoughts, either related or unrelated to

the shooting. They made these ratings on a 6-point Likert scale (1= not at all, 6= completely). Subjects

in the no witness condition were asked to report only whether they thought the husband shot his wife.

Finally, as another attention check, subjects saw an open text box and were asked to write who [the

witness] was in relation to the husband.

Results and Discussion

Recall the primary purpose of this experiment was to answer three research questions. First, to

what extent do people believe a parrot is a credible witness compared to a three-year-old child?

Second, to what extent are people persuaded by an utterance made by a parrot when making

judgements of guilt compared to a child? Finally, to what extent do people rate a parrot's utterance as

a unique thought or repetition compared to a three-year-old child? Before addressing our research questions, we first excluded 18 subjects who failed the attention check, were incorrect about who Leroy (the witness) was, or were inconsistent on the reversed questions. For example, responding 5 on the item "[witness]'s account of events can be depended upon" and also responding 5 on the item "Leroy's memory of the event is likely to be unreliable" would be considered inconsistent. Our final sample consisted of 306 subjects.

Credibility. To what extent did people believe a parrot is a credible witness compared to a three-year-old child? To answer this question, we calculated the mean score of each subscale item in the CSAVCS. We found in Experiment 1 that subjects in the parrot condition slightly disagreed with the credibility statements but did not dismiss the parrot entirely. One possible explanation was that the truthfulness subscale was inflating the overall credibility mean as people may believe that a parrot is unable to lie and is ultimately truthful. We therefore needed to check there was a consistent pattern across each individual subscale of the Child Sexual Assault Victim Credibility Scale (CSAVCS). To do this, we ran a Hotelling's T^2 as this allowed us to compare multiple variables across a single measure between groups. As shown in Table 2, across all five subscales, we found that a three-year-old child was rated as more credible than an African grey parrot.

In NHST terms, we ran a between-subjects Hotelling's two-sample T^2 test with witness type (child, parrot) as a between-subject factor and the credibility subscales as dependent variables. We found a significant difference between the parrot and child conditions; $T^2(5,208)= 7.38$, $p < .001$. We then conducted a post hoc between samples t-test with a Bonferroni correction on each subscale and, as seen in Table 2, found that the child was rated as more accurate, believable, competent, reliable, and truthful compared to the parrot. In other words, the results indicated that the mean credibility ratings for the parrot condition were significantly different than the child condition across all five subscales. These results replicate our finding in Experiment 1 that people rated parrots as less credible than three-year-old children but did not dismiss the parrot's evidence entirely.

Table 2

Mean, 95% Confidence Interval and Standard Deviation of each Subscale across the Witness Conditions (parrot, child)

with Mean Difference and 95% Corrected Confidence Intervals of Mean Difference for Multiple Analyses

	Parrot			Child			Mean difference	
Subscale	M	95% CI	SD	M	95% CI	SD	M_{diff}	95% CI
Accuracy	2.66	[2.41, 2.91]	1.27	3.64	[3.40, 3.88]	1.19	0.98	[0.52, 1.43]
Believability	3.03	[2.75, 3.32]	1.44	3.98	[3.74, 4.21]	1.16	0.94	[0.46, 1.43]
Competency	2.79	[2.52, 3.05]	1.35	3.46	[3.21, 3.71]	1.25	0.67	[0.19, 1.16]
Reliability	2.51	[2.22, 2.80]	1.46	3.62	[3.35, 3.90]	1.35	1.11	[0.59, 1.63]
Truthfulness	3.07	[2.78, 3.35]	1.43	4.20	[3.96, 4.44]	1.18	1.13	[0.65, 1.62]

Note. Reported mean difference confidence intervals are 95% confidence intervals corrected for multiple comparisons.

Guilt. We now turn to our second research question: To what extent do people rate a suspect as guilty based on a parrot or three-year-old child's utterance compared to no utterance? To address this question, we classified the guilt ratings for the husband according to whether the subject is in the parrot, child, or no witness condition, and displayed the data in Figure 2. As the figure shows, we found that how subjects rated the perceived guilt of the husband depended on the type of "witness" who uttered the phrase "[husband] did it." In other words, subjects believed the husband was more likely to be guilty when they were informed that the child was repeating the phrase "[husband] did it" compared to subjects who were not told about a witness.

In statistical terms, a one-way between-subjects ANOVA showed an effect of witness on husband's guilt rating, $F_{(2,303)} = 4.30$, $p = 0.014$, $\eta2 = .028$ 90% $\eta2$ CI $[0.003, 0.060]$. Post hoc comparisons using the Turkey HSD test showed that the mean guilt rating for the child condition was significantly higher than for the no witness condition ($M_{diff} = 0.74$, 95% CI $[0.09, 0.85]$). But the rating for the parrot condition did not significantly differ either from the child ($M_{diff} = 0.26$, 95% CI $[-0.13, 0.64]$), or from the no witness conditions ($M_{diff} = 0.21$, 95% CI $[-0.16, 0.58]$).

Figure 2

Mean Guilt Rating across Witness Conditions (control, parrot, child)
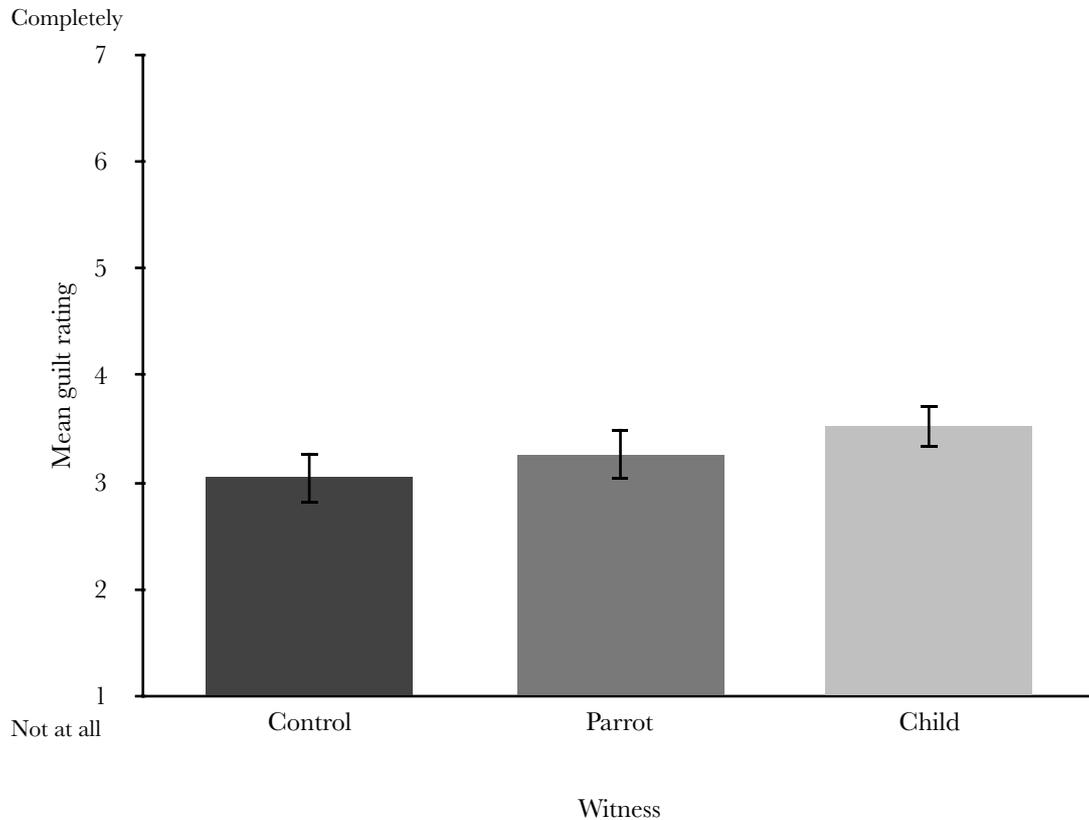
Note. Error bars represent 95% confidence intervals of cell means.

Interpretation of utterance. To address our third question about the extent to which people rate a parrot's utterance as a unique thought or repetition compared to the three-year-old child, we calculated the mean score for each utterance item according to whether the subject is in the parrot or child condition. We then ran independent sample t-tests on each interpretation of the utterance item with a Bonferroni correction. As Figure 3 shows, subjects were more likely to believe the parrot was repeating an utterance unrelated to the crime compared to the child. Additionally, subjects were more likely to believe the child was producing a unique utterance related to the crime compared to the parrot.

In statistical terms, four independent-samples t-tests with a Bonferroni correction were conducted to compare how people interpret the child or parrot's utterance. As seen in Figure 3, there

was a significant difference in the "[Witness] is attempting to communicate his own unique thoughts related to the murder" item between the parrot condition and the child condition (M_{diff} = 1.41, 95% $CI_{corrected}$ [0.90, 1.93]). There was also a significant difference in the "[Witness] is repeating something he heard unrelated to the murder" item between the parrot and the child condition, (M_{diff} = 1.21, 95% $CI_{corrected}$ [-1.71, -0.71]). We did not find a significant difference in the "[Witness] is attempting to communicate his own unique thoughts unrelated to the murder" item between the parrot condition and the child condition (M_{diff} = 0.03, 95% $CI_{corrected}$ [-0.58, 0.51]). Nor was there a significant difference in the "[Witness] is repeating something he heard related to the murder" item between the parrot condition and the child condition (M_{diff} = 0.173, 95% $CI_{corrected}$ [-0.72, 0.37]).

Figure 3

Mean Item Ratings across Interpretation of Utterance Items by Witness Conditions (parrot, child)
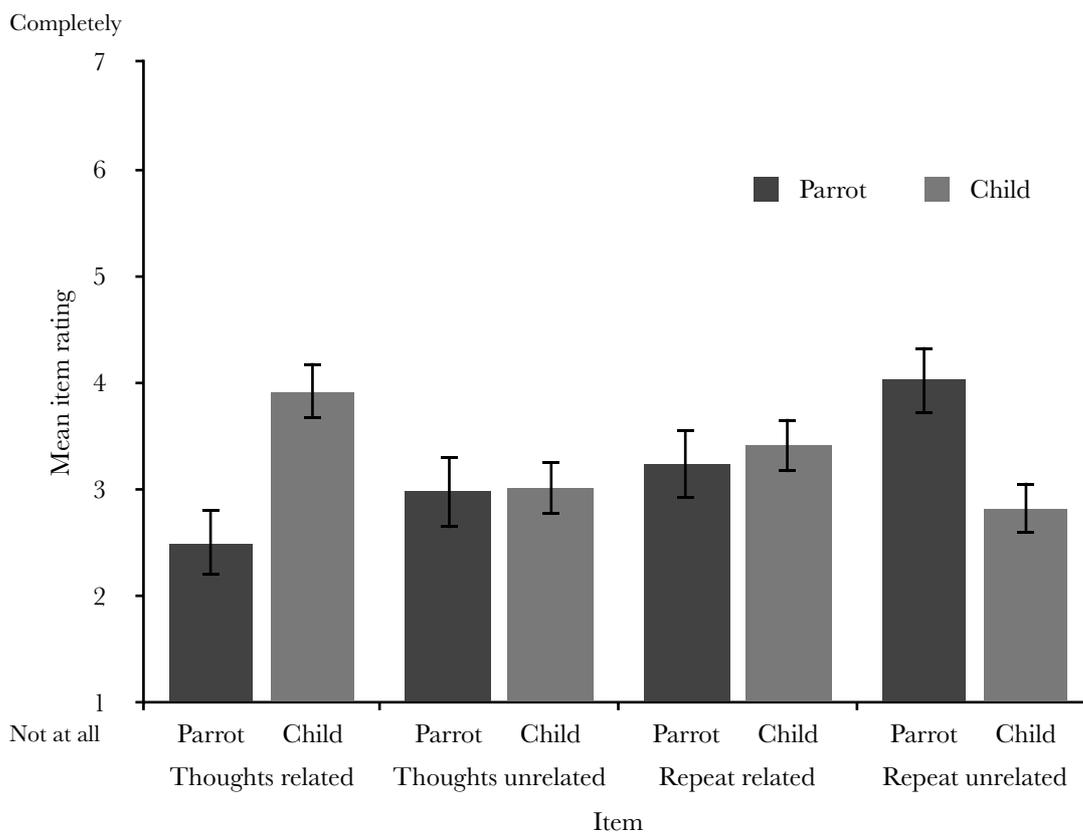
Note. Error bars represent 95% confidence intervals of cell means.

Experiment 2 produced similar results to Experiment 1. We found that subjects rated a parrot as less credible than a child, but still did not dismiss the parrot's evidence entirely. We also obtained some evidence to counter the idea that the CSAVCS anchors led subjects to respond to each item differently than they otherwise might have, had the scale measured credibility. We also obtained evidence that the truthfulness subscale was not responsible for inflating the mean credibility score. Additionally, we were able to dismiss the possibility that subjects were mistakenly believing that the parrot was producing unique thoughts, rather than repeating something overheard. Experiment 2 demonstrated that subjects believe that the parrot is more likely to be repeating an utterance unrelated to the crime compared to the chid. Taken together our data fit with the idea that people believe the parrot is actually somewhat credible.

We also found that people may be inclined to consider the parrot's utterance when making judgements of guilt. We found that subjects in the parrot condition did not rate the husband's guilt as significantly higher or lower than subjects in the child or no witness condition. But there is evidence suggestive of a linear trend that would be more obvious with a larger sample size which would allow for a more precise estimate of the true size of the effect (Cumming, 2014). Of course, the obvious counter-explanation is that there is no actual difference in the population. We further address these issues in Experiment 3.

Experiment 3

Experiments 1 and 2 demonstrated that people rate a parrot as less credible than a three-year-old, but do not dismiss the parrot's evidence entirely. We know that people often struggle to recognise the ambiguity of statements (Wylie et al., 2019). People are also less critical of information that fits with their beliefs (Jonas et al., 2001; Nickerson, 1998; Russo, 2014). If people believe the husband committed the crime, then they are likely to be less critical of information that supports that belief. Therefore, we hypothesise that people will be more likely to be persuaded by ambiguous utterances compared to no utterance when making judgments of guilt. In Experiment 3, we asked three research questions. First, to what extent do people rate a parrot as a credible witness when the parrot repeats a

highly ambiguous utterance compared to a less ambiguous utterance? Second, to what extent would people be persuaded by a highly ambiguous utterance compared to a less ambiguous utterance or no utterance when making judgements of guilt? Finally, to what extent do people believe the parrot's utterance is a repetition of an overheard phrase and to what extent do people believe that same utterance relates to the crime? We preregistered this experiment (see Appendix C for the pre-registration).

Method

Subjects. Subjects were recruited via Amazon's Mechanical Turk. A total of 663 US Mechanical Turk workers completed the experiment in exchange for 0.25 USD. The subjects' age ranged from 18 to 87 years ($M = 41.21$, 95% CI [40.18, 42.24], $SD = 13.56$); 403 subjects identified as women, 259 identified as men, 1 identified as non-binary; 96.39% of subjects reported English as their first language, and 99.70% reported English as their primary language, the language they use most day-to-day.

Design. We used a simple three-condition design with the ambiguity of the utterance (high ambiguity, low ambiguity, no utterance) manipulated between subjects.

Procedure. Subjects completed the experiment online via Qualtrics. Before beginning the study, subjects provided their consent. We again told subjects that the study was examining learning styles to ensure that subjects were unaware of the true purpose of the study. We asked that subjects completed the experiment in lab-like conditions.

Subjects read the same vignette used in Experiments 1 and 2; however, there was one notable change. The vignette subjects read was changed to inform the subjects that "a neighbor reported that [husband's name] and his wife had been fighting the night before." The purpose of this statement was to highlight the potential ambiguity of the utterance "[husband] don't" as this phrase could make sense in many different circumstances, including in the context of an argument. This additional piece of information was added to subtly draw attention to the fact that the ambiguous statements were in fact, ambiguous.

Subjects were then told the husband's African grey parrot had repeatedly been saying an utterance in the days following the death. Subjects in the high ambiguity condition were told that the utterance was "[husband] don't." Subjects in the low ambiguity condition were told that the utterance was "[husband] don't shoot." Subjects in the no utterance condition were not told about an utterance but were informed that [husband]'s pet African grey parrot was inside the house the day the death occurred.

We measured three different judgements of guilt. Subjects were first asked to identify if they believe the husband was guilty or not guilty: "With regard to the crime of Mrs. [husband's last name]'s murder, I think that Mr. [husbands name] is" (guilty, not guilty). The screen position in which the options guilty and not guilty were presented was counterbalanced. The subjects were then asked to rate their confidence in their judgement of judicial guilt on a 7-point Likert scale (1= not at all confident, 7= very confident). Additionally, subjects were asked to rate the likelihood of the husband having shot his wife on a 7-point Likert scale (1= very unlikely, 7= very likely). Subjects in the high and low ambiguity conditions were asked to then complete the adapted Child Sexual Assault Victim Credibility Scale (CSAVCS) used in Experiment 2 about the parrot. They were also asked two questions similar to Experiment 2 regarding how they interpreted the utterance: "I believe [the witness]'s utterance is something he overheard" and "I believe [the witness]'s utterance is related to the death of Mrs. [husband's last name]" on a 7-point Likert scale (1= not at all, 7= completely).

Results and Discussion

Before addressing our research question, we first excluded 72 subjects who failed to complete the survey, failed either of the attention checks, or were incorrect about who Leroy was. Our final sample consisted of 593 subjects. We diverted from our pre-registration when analysing people's confidence in guilt. We had pre-registered to conduct planned comparisons between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' ratings of confidence of guilt "How confident are you in your judgement that Mr. [first name] [last name] is [guilty/not guilty]". Instead we split the the conditions by subjects' judicial guilt selection (guilty, not guilty) as this allows us

to investigate whether subjects are more confident when they believe the husband is guilty compared to when they believe the husband is not guilty.

Credibility. To what extent did people rate a parrot as a credible witness when the parrot repeats a highly ambiguous utterance compared to a less ambiguous utterance, To answer this question, we calculated the mean of the 21 credibility items from the CSAVCS by utterance condition, classified them according to whether subjects read an utterance that was high ambiguity, low ambiguity, or they read no utterance, and displayed the results in Figure 4. Because we found that in Experiment 2, across all five subscales, the parrot was rated less credible than the child (see Table 2), we collapsed the subscales to measure overall mean credibility. As the figure shows, subjects rated the witness as more credible when the utterance was less ambiguous compared to a highly ambiguous utterance. In other words, subjects believed the witness was more credible when the utterance was "[husband] don't shoot" compared to "[husband] don't" ($M_{\text{diff}} = 0.28$, 95% CI [0.02, 0.54]).

Judicial Guilt. We now turn to our second question: To what extent would people be persuaded by a highly ambiguous utterance compared to a less ambiguous utterance or no utterance when making judgements of guilt? To answer this question, we first calculated the proportion of guilty verdicts by utterance condition, "[husband] don't shoot", [husband] don't", and no utterance, and displayed the data in Figure 5.

As the figure shows, we found that more subjects believed the husband was guilty when they were informed about an utterance compared to subjects who did not receive an utterance. More subjects who were informed that the utterance was "[husband] don't shoot" believed the husband was guilty than subjects who were informed the utterance was "[husband] don't".

In statistical terms, following the proportion estimation approach (Cumming & Calin-Jageman, 2017) we conducted a difference in proportions comparison between each of the utterance conditions on whether subjects believe the husband was guilty or not guilty. We found that subjects were more likely to rate the husband as guilty when the witness (an African grey parrot) was repeating an utterance. We found this for both levels of ambiguity, which is to say subjects were more likely to rate the husband as guilty both when the utterance was "[husband] don't" and "[husband] don't shoot."

Figure 4

Mean Credibility Rating of 21 Credibility Items by Utterance Condition (less ambiguous "[husband] don't shoot",

highly ambiguous "[husband] don't")
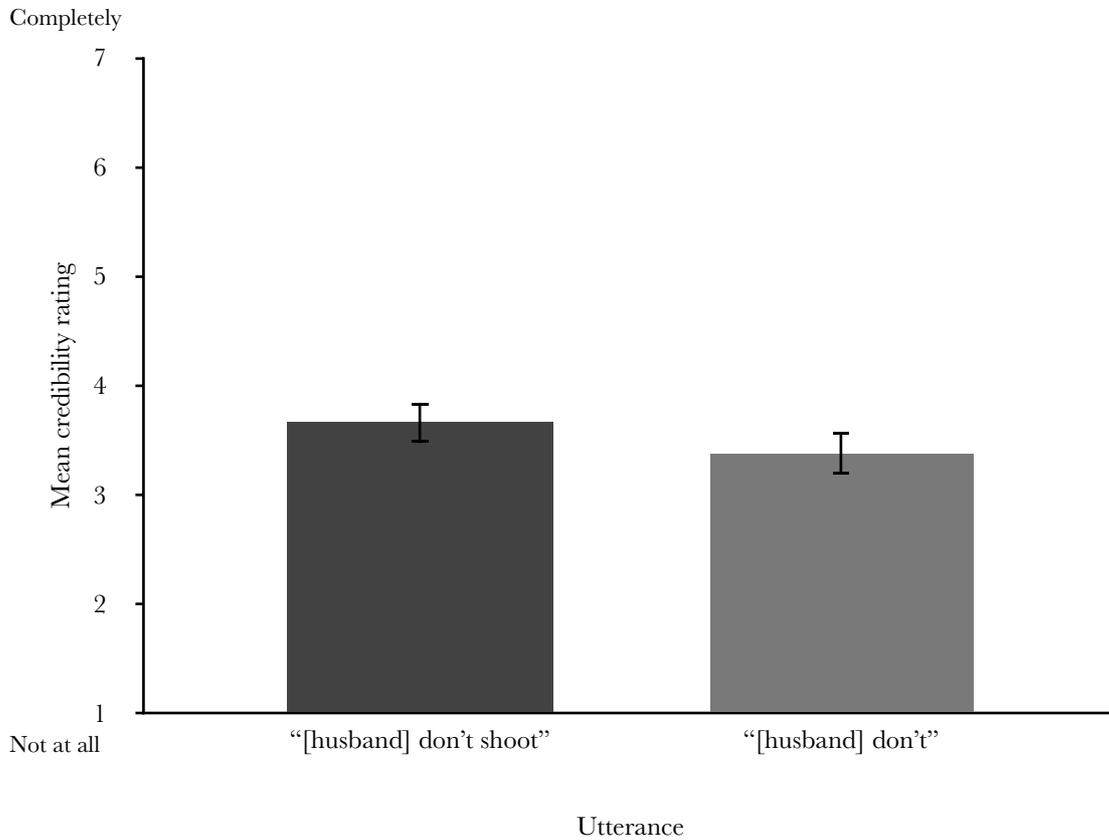
Note. Error bars represent 95% confidence intervals of cell means.

As shown in Figure 5, a higher proportion of subjects judge the husband as guilty both when the utterance was "[husband] don't shoot" ($M_{\text{diff}} = 0.32$, 95% CI [0.22, 0.41]) and "[husband] don't" ($M_{\text{diff}} = 0.14$, 95% CI [0.04, 0.23]) compared to no utterance. Additionally, a higher proportion of subjects judged the husband as guilty when the utterance was "[husband] don't shoot" compared to when the utterance was "[husband] don't" ($M_{\text{diff}} = 0.18$, 95% CI [0.08,0.27]).

Confidence in guilt verdict. To address our research question: To what extent would people be persuaded by a highly ambiguous utterance compared to a less ambiguous utterance or no utterance when making judgements of guilt? We also need to consider how confident people are in their guilt verdict. Therefore, we calculated subjects' mean ratings of confidence in guilt verdict by utterance type and whether the subject believed the husband was guilty or not guilty.

Figure 5

Proportion of Guilty Verdicts across the Utterance Conditions (less ambiguous "[husband] don't shoot", highly ambiguous "[husband] don't", no utterance)
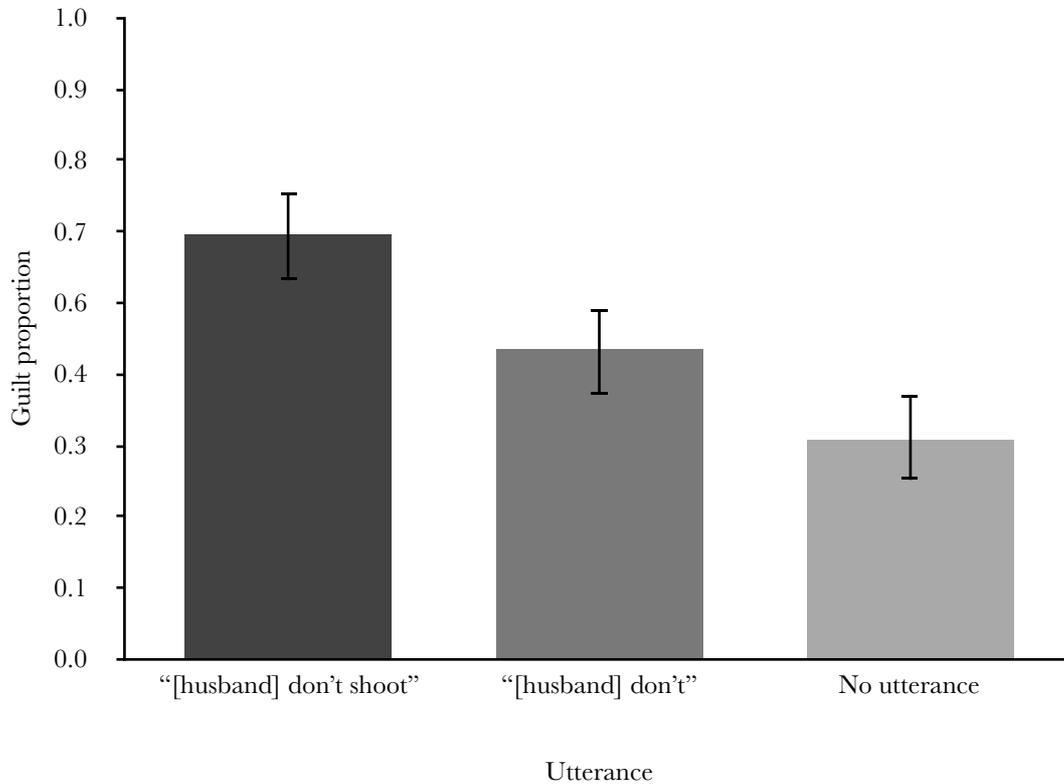
Note. Error bars represent 95% confidence intervals of condition proportions.

As Figure 6 shows, we found that subjects who believed the husband was guilty were more confident when the utterance was either "[husband] don't shoot" or "[husband] don't" compared to subjects who believed the husband was not guilty. We did not find any difference between subjects who believed the husband was guilty when there was no utterance compared to subjects who believed the husband was not guilty.

In statistical terms, we calculated mean differences between guilty and not guilty ratings across the three ambiguity conditions. As seen in Figure 6, subjects who rated the husband as guilty were more confident in their verdict when told the parrot was repeating the utterance "[husband] don't" compared to subjects who rated the husband as not guilty ($M_{diff} = 0.91$, 95% CI [0.52, 1.29]). Similarly, subjects who rated the husband as guilty were more confident in their verdict when told the

parrot was repeating the utterance "[husband] don't shoot" compared to subjects who rated the husband as not guilty (M_{diff} = *0.70,* 95% CI [0.28, 1.11]). We did not find any difference between subjects who rated the husband as guilty compared to not guilty when no utterance was provided (M_{diff} = *0.17,* 95% CI [-0.32, 0.65]).

Likelihood of guilt. To address the research question: To what extent would people be persuaded by a highly ambiguous utterance compared to a less ambiguous utterance or no utterance when making judgements of guilt? We also needed to investigate how likely people thought it was that the husband murdered his wife. To address this question we calculated subjects' mean likelihood rating across utterance conditions and displayed the data in Figure 7.

Figure 6

Mean Confidence in Guilt Rating across the Utterance Conditions (less ambiguous "[husband] don't shoot", highly ambiguous "[husband] don't", no utterance) by Guilt Verdict
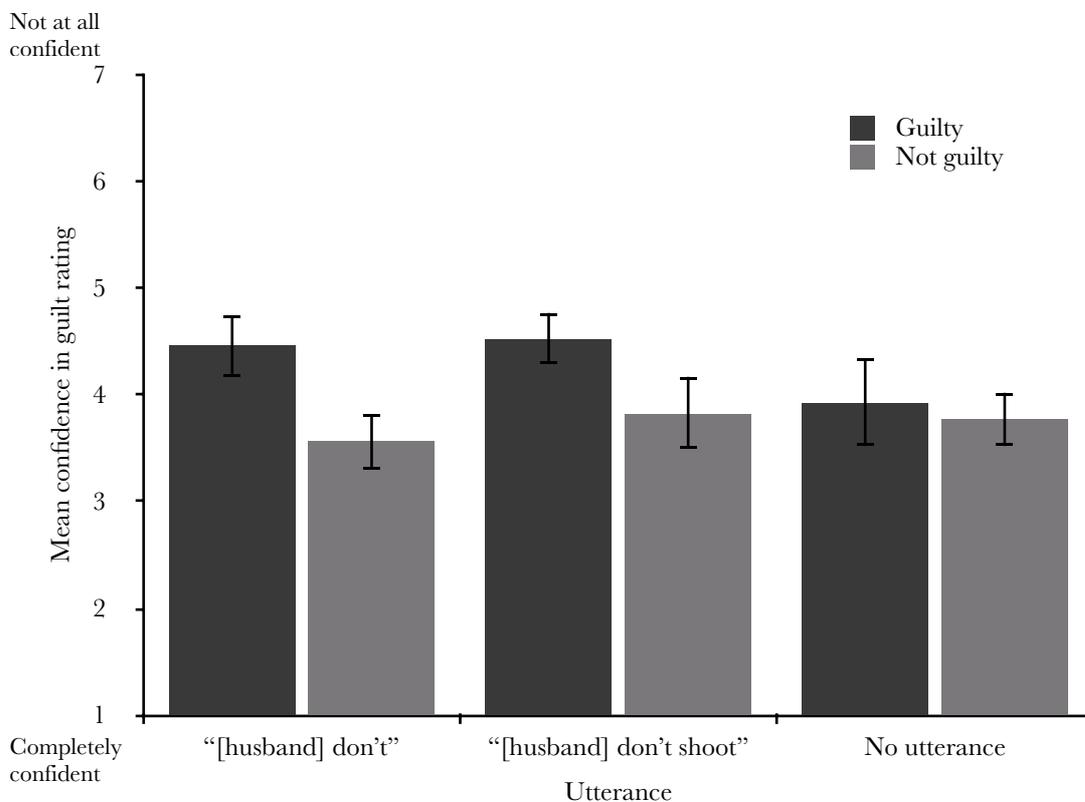
Note. Error bars represent 95% confidence intervals of cell means.

As the figure below shows, we found that subjects rated the likelihood of the husband shooting his wife higher when the witness provided an utterance, regardless of the level of ambiguity, compared to no utterance. Additionally, we found that subjects rated the likelihood higher when the utterance was less ambiguous.

In statistical terms, we calculated the mean differences between the three ambiguity conditions and subjects' rating of likelihood of guilt. As seen in Figure 7, subjects rated the husband as more likely to be guilty when the utterance was "[husband] don't shoot" compared to "[husband] don't" (M_{diff} = 0.26, 95% CI [0.02, 0.50]). Similarly, subjects rated the husband as more likely to be guilty when the utterance was "[husband] don't shoot" compared to no utterance (M_{diff} = 0.59, 95% CI [0.35, 0.83]). Additionally, subjects rated the husband as more likely to be guilty when the utterance was "[husband] don't shoot" compared to "[husband] don't" (M_{diff} = 0.33, 95% CI [0.10, 0.57]).

Interpretation of utterance. To address our final research question: To what extent do people believe the parrot's utterance is a repetition of an overheard phrase, and to what extent do people believe that same utterance relates to the crime? We calculated the mean scores of the two utterance interpretation items "I believe the [Witness]'s utterance is related to the murder" and "I believe the [Witness]'s utterance is something he overheard" by utterance "[husband] don't shoot" or "[husband] don't." We found that when subjects were informed that the utterance was "[husband] don't shoot", subjects believed the utterance was more likely to be related to the crime and a repetition of something overheard compared to subjects who were informed the utterance was "[husband] don't."

In statistical terms, we conducted a planned comparison between the two utterance conditions on subjects' interpretation of utterance items. We found that when the utterance was "[husband] don't shoot" subjects were more likely to rate the witness's utterance as being related to the murder (M_{diff} = 0.78, 95% CI [0.45, 1.12]) and something overheard (M_{diff} = 0.36, 95% CI [0.06, 0.65]) compared to "[husband] don't."

Figure 7

Mean Likelihood Rating Across Utterance Condition (less ambiguous "[husband] don't shoot", highly ambiguous "[husband] don't", no utterance)
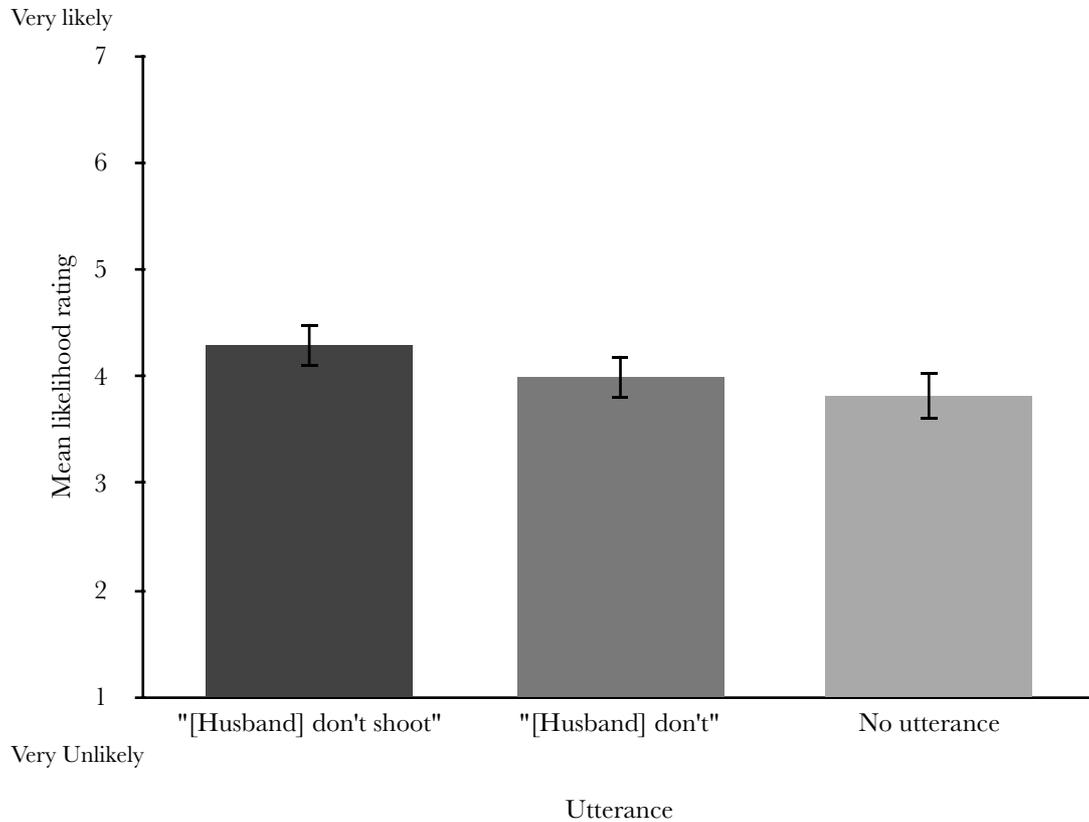
Note. Error bars represent 95% confidence intervals of cell means.

In this experiment, we replicated the finding that the parrot was not rated as all that credible but was not dismissed entirely. What is interesting about these results is that we found the parrot's rating of credibility differed between the utterance "[husband] don't shoot" and "[husband] don't". A witness's credibility should not depend on the utterance. In an ideal world, people would assess the credibility of the witness before considering the information from that witness. Instead, we found that the utterance informed the credibility of the witness, which suggests that in addition to non-credible related cues like confidence, people are also using the witness's statement to assess credibility. The problem of using a witness's statement to inform credibility is that people will be more likely to accept information that fits with their story of events (Pennington & Hastie, 1991, 1992). This sets up a

"begging the question" fallacy where people believe the witness is more credible because the statement fits their story and that their story is correct because the witness is credible.

We also found subjects were more likely to rate the husband as guilty when presented with either the less ambiguous utterance "[husband] don't shoot" or the highly ambiguous utterance "[husband] don't" compared to no utterance. Both the highly ambiguous statement "[husband] don't", and the less ambiguous "[husband] don't shoot" increased guilty verdicts. These results suggest that people can be persuaded with highly ambiguous information from less-than-credible sources when making judgments of guilt. Additionally, people were more confident in their guilty verdicts when presented an utterance compared to those who believed the husband was innocent. These results suggest that people who believed the husband was guilty were using the utterance to support their decision, regardless of the ambiguity of the utterance. These results were supported by the likelihood of guilt item. We found a linear trend that suggests people were more likely to be persuaded that the husband murdered his wife when the parrot provided an utterance compared to people who did not receive an utterance. The finding that people are more persuaded by an ambiguous utterance from less-than-credible sources compared to no utterance supports our hypothesis. Although the effect is small, these results are still meaningful as even a slight increase can mean the difference between a guilty and innocent verdict and could be the difference between an innocent suspect spending life in prison or going free.

General Discussion

Across three experiments and 1191 subjects, we found evidence that although people did not believe parrots are all that credible compared to a three-year-old, they did not dismiss a parrot's credibility entirely. Despite finding that parrots are not all that credible, we found evidence that people can be persuaded by an utterance from a parrot when making judgements of guilt, even when the utterance is ambiguous.

In Experiment 1, we demonstrated that people will rate a parrot as less credible than a three-year-old child, but do not dismiss the parrot entirely. This finding was further expanded upon in Experiment 2, which again demonstrates that people rated the parrot as less credible, but still believed

the parrot was somewhat credible (M= 2.89, 95% CI [2.63, 3.14]). Experiment 2 also suggested a linear trend in that subjects who were informed that the parrot was repeating the utterance "[husband] did it" believed the husband was more likely to be guilty than subjects who were not informed about a witness; however, this difference was not statistically significant. The linear trend suggests a larger sample size would allow for a more precise estimate of the true size of the effect (Cumming, 2014).

Experiment 2 also demonstrated that people believed the parrot's utterance "[husband] did it" was most likely to be a repetition unrelated to the crime. Given that people believed the utterance was a repetition, we adapted the utterance in Experiment 3 to a plausible phrase that could have been overheard. With these new utterances, Experiment 3 demonstrated that people can be persuaded by ambiguous utterances from less-than-credible sources, when making a judgement of guilt. Furthermore, we demonstrated that people, when presented with the parrot's utterance, were more confident in their verdict when they believed that the husband was guilty, compared to people who believed the husband was not guilty. One explanation is that the utterance fits their belief that the husband murdered his wife. People's increased confidence when rating the husband guilty suggests that people are in fact being persuaded by the parrot's utterance.

Of course, there are at least three obvious caveats to these conclusions. First, across all three experiments, subjects were provided with only two pieces of information: The vignette and limited information about the witness. It is, therefore, plausible that our observed effect is reliant upon no other information being available. In a typical case, a jury would be presented with multiple pieces of information to consider before making a judgement of guilt. Because our subjects were presented with a very small amount of information, that information might have had more impact that it would have otherwise if more credible sources were available. Finally, it is also possible that subjects were encouraged to interpret the utterance as evidence, because the vignette told the subjects that the police had no evidence; but now have new information and want to consider it as evidence. As the police wanted to use the information as evidence, the subjects may have believed that the utterance should be considered as evidence. But the possibility that people were encouraged to believe the information is

evidence is a problem in real-world courtrooms as well. Individual factors such as personal experience and the situation in which people encounter ambiguous information can encourage people to interpret ambiguous information in a certain way (Duffy & Feist, 2014). Given that people would likely expect information provided in a courtroom to be evidence, we would expect that people would be encouraged to interpret ambiguous information as evidence when in a courtroom. Future research should focus on investigating the possible alternative explanation that the effect we found only exists because no other information was provided. In a courtroom, jurors will often review multiple pieces of information before being asked to make a judgement of guilt. Ideally, a future study would manipulate the amount of information provided in the overall case to establish whether the effect observed in our experiments still hold when more credible sources are available.

The second caveat is that our experiments focused on a murder case. People often feel more pressure to resolve murders than less severe crimes and may have a bias to obtain a guilty verdict if given a reason to do so (Lundrigan et al., 2018; Wiener et al., 2006). Therefore, people may be willing to look more broadly for evidence and information to inform their judgement of guilt than they would typically when dealing with a less serious case. It is plausible that the effect observed in our experiments may reduce or fail to replicate in cases involving less serious crimes such as fraud and other white-collar crimes. Alternatively, we might plausibly see the opposite effect, in that people might engage in more systematic decision making when working on a higher-pressure case. We know that people are less likely to consider inadmissible evidence when judging more serious cases compared to less serious cases (Rind et al., 1995). Therefore, we may see people are more willing to consider an ambiguous utterance from a less-than-credible witness in cases involving less serious crimes. Regardless of the direction of the effect, future research could further investigate whether the effect replicates in less serious cases by manipulating the severity of the crime.

Finally, a critic might be concerned that we did not norm the ambiguity of our utterances. Although we are confident that "[husband] don't" is a more ambiguous utterance than "[husband] don't shoot",we acknowledge that a critic could argue that the utterance "[husband] don't shoot" is not ambiguous and that there are few opportunities in which a parrot would overhear this phrase, and

therefore, we are not really manipulating ambiguity. But our finding observed in Experiment 3, that people can be persuaded by ambiguous statements from less-than-credible sources, does not support that idea as we found people were still persuaded by the highly ambiguous utterance "[husband] don't" compared to the control condition.

These findings have implications for work in other literatures. First, our finding that people can be persuaded by ambiguous statements from less-than-credible sources extends our current knowledge of how people interpret ambiguous information. The literature on ambiguous information suggests that people often fail to recognise ambiguity, and that people will interpret ambiguous statements through their own experiences and beliefs (Dearing & Gotlib, 2009; Lau et al., 2011; Wylie et al., 2019). Our finding provides evidence that people will interpret and be persuaded by ambiguous statements regardless of the credibility of the source. This finding also expands on the body of research which suggests that people will be persuaded by information that they objectively should not such as inadmissible evidence and pretrial publicity (Daftary-Kapur et al., 2010; Daftary-Kapur et al., 2014; Kaufmann et al., 2003; Nitschke et al., 2019; Ruva et al., 2007; Ruva & Guenther, 2015, 2017; Schuller et al., 2010; Steblay et al., 2006; Steblay et al., 1999). Our finding shows that people can also be influenced by information from a source that they do not consider credible.

Second, our finding that people are more confident of their judgement when a less-than-credible witness supports their belief is consistent with the Story Model and predecisional bias literature. We know from the Story Model that people will construct their own stories when assessing information and judge which story is most likely to be correct (Pennington & Hastie, 1992, 1991). We also know from the predecisional bias literature that people are less critical of information that fits with their preferred story (DeKay, 2015; Russo, 2014; Taber & Lodge, 2006). Our finding supports that literature as we found evidence that people who constructed a story of murder in which the husband was guilty were more confident in their decision when the parrot was repeating an ambiguous utterance. The increased confidence in judgement suggests that people's confidence was bolstered by the parrot's utterance, despite the fact that the utterance was ambiguous and from a less-

than-credible source. Therefore, it is plausible that the reason people were less critical of the information is because it supported their story that the husband murdered his wife.

Finally, we found that people rated the parrot as more credible when the utterance was less ambiguous. This finding expands upon the a growing body of literature which suggests that people will consider non-crediblity related cues when assessing credibility (Jiang et al., 2020; Kaufmann et al., 2003; McCarthy Wilcox & NicDaeid, 2018; Nitschke et al., 2019; Price & Stone, 2004; Sah et al., 2013). We have evidence that people will consider a witness's statement when judging credibility but this logic can lead to a circular argument, in which people believe that the parrot is more credible because the utterance fits their story, and therefore, their story is correct because the parrot is a credible witness.

Taken together, our and prior work suggests that courts need to be especially careful about what information is provided in a courtroom. If a witness provides an ambiguous response or statement in court, we need to ensure that the ambiguity of the statement is removed through further probing as people may assume the statement means one thing without realising that the statement is actually ambiguous and could mean another thing. We cannot expect jurors to monitor what information should and should not be used when making judgements of guilt. Ideally, we would ensure that the information being provided to the jury is clear and concise information from objectively credible sources. Obviously, this would not be practical, but regardless we need to consider what information is provided to the jury as we found that people may consider ambiguous information from less-than-credible sources.

Our research also has theoretical implications for our understanding of the relationship between credibility and judgements of guilt. Prior work suggests that people will first assess credibility to inform whether or not they will actually consider that source's information (Brodsky et al., 2010; Goodman-Delahunty et al., 2010; Voogt & Klettke, 2017; Voogt et al., 2020). But our research suggests that in the absence of credible sources, people will be persuaded by information from less-than-credible sources, even when the information is ambiguous. Jurors often are the final guard to ensure that the innocent are not unjustifiably charged with crimes they did not commit and that the guilty are held

accountable for their actions. We, therefore, need to ensure that we are not unintentionally corrupting

that system by allowing less-than-credible sources or ambiguous information to inform jurors'

judgements. After all, we should never send someone to prison just because a little birdie told us to.

References

Abramowitz, J. S., Tolin, D. F., & Street, G. P. (2001). Paradoxical effects of thought suppression: a meta-analysis of controlled studies. *Clinical Psychology Review, 21*(5), 683-703. https://doi.org/https://doi.org/10.1016/S0272-7358(00)00057-X

Bottoms, B. L., Davis, S. L., & Epstein, M. A. (2004). Effects of victim and defendant race on jurors' decisions in child sexual abuse cases. *Journal of Applied Social Psychology, 34*(1), 1-33. https://doi.org/10.1111/j.1559-1816.2004.tb02535.x

Bottoms, B. L., Goodman, G. S., Schwartz-Kenney, B. M., & Thomas, S. N. (2002). Understanding children's use of secrecy in the context of eyewitness reports. *Law and Human Behavior, 26*(3), 285-313. https://doi.org/10.1023/A:1015324304975

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law, 28*(6), 892-907. https://doi.org/10.1002/bsl.917

Ceci, S. J., & Bruck, M. (2006). Children's suggestibility: Characteristics and mechanisms. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 247-281). Academic Press. https://doi.org/10.1016/S0065-2407(06)80009-1

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. https://doi.org/10.1007/BF02310555

Cumming, G. (2014). The new statistics:Why and how. *Psychological Science*, 25(1), 7-29. https://doi.org/10.1177/0956797613504966

Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the New Statistics* (1 ed.). Routledge. https://doi.org/10.4324/9781315708607

Daftary-Kapur, T., Dumas, R., & Penrod, S. D. (2010). Jury decision-making biases and methods to counter them. *Legal and Criminological Psychology, 15*(1), 133-154. https://doi.org/10.1348/135532509x465624

Daftary-Kapur, T., Penrod, S. D., O'Connor, M., & Wallace, B. (2014). Examining pretrial publicity in a shadow jury paradigm: Issues of slant, quantity, persistence and generalizability. *Law and Human Behavior, 38*(5), 462-477. https://doi.org/10.1037/lhb0000081

Dearing, K. F., & Gotlib, I. H. (2009). Interpretation of ambiguous information in girls at risk for depression. *Journal of Abnormal Child Psychology, 37*(1), 79-91. https://doi.org/10.1007/s10802-008-9259-z

DeKay, M. L. (2015). Predecisional information distortion and the self-fulfilling prophecy of early preferences in choice. *Current Directions in Psychological Science, 24*(5), 405-411. https://doi.org/10.1177/0963721415587876

Demaine, L. J. (2008). In search of the anti-elephant: Confronting the human inability to forget inadmissible evidence. *George Mason Law Review, 16*, 99-140. https://heinonline.org/HOL/Page?collection=journals&handle=hein.journals/gmlr16&id=100&men_tab=srchresults

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior, 19*(6), 695-704. https://doi.org/10.1016/S0022-5371(80)90379-5

Duffy, S., & Feist, M. (2014). Individual differences in the interpretation of ambiguous statements about time. *Cognitive Linguistics, 25*, 29-54. https://doi.org/10.1515/cog-2013-0030

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition, 33*(5), 770-782. https://doi.org/10.3758/BF03193073

French, L., Garry, M., & Mori, K. (2011). Relative – not absolute – judgments of credibility affect susceptibility to misinformation conveyed during discussion. *Acta Psychologica, 136*(1), 119-128. https://doi.org/https://doi.org/10.1016/j.actpsy.2010.10.009

Golding, J. M., Wasarhaley, N. E., Lynch, K. R., Lippert, A., & Magyarics, C. L. (2015). Improving the credibility of child sexual assault victims in court: The impact of a sexual assault nurse examiner. *Behavioral Sciences & the Law, 33*(4), 493-507. https://doi.org/10.1002/bsl.2188

Goodman-Delahunty, J., Cossins, A., & O'Brien, K. (2010). Enhancing the credibility of complainants in child sexual assault trials: The effect of expert evidence and judicial directions. *Behavioral Sciences & the Law, 28*(6), 769-783. https://doi.org/10.1002/bsl.936

Hackett, L., Day, A., & Mohr, P. (2008). Expectancy violation and perceptions of rape victim credibility. *Legal and Criminological Psychology, 13*(2), 323-334. https://doi.org/10.1348/135532507x228458

Hanlon, R. E., Brook, M., Demery, J. A., & Cunningham, M. D. (2016). Domestic homicide: Neuropsychological profiles of murderers who kill family members and intimate partners. *J Forensic Sci, 61 Suppl 1*, S163-170. https://doi.org/10.1111/1556-4029.12908

Harris/Decima Inc. (2010, December 15). *More than half of Canadians trust their pets more than people* [Press release]. Nestle Purina PetCare. Retreived from https://www.newswire.ca/news-releases/more-than-half-of-canadians-trust-their-pets-more-than-people-507184761.html

Harris Interactive. (2012, September 13). Pets aren't just animals; They are members of the family. https://theharrispoll.com/new-york-n-y-september-13-2012-americans-have-always-had-interesting-relationships-with-their-pets-whether-those-pets-are-cats-dogs-parakeets-or-something-else-entirely-the-pet-industry-i/

India Today Online. (2014). How a parrot helped Agra cops crack a difficult murder case. *India Today*. https://www.indiatoday.in/india/north/story/how-a-parrot-helped-agra-cops-crack-a-difficult-murder-case-182812-2014-02-26

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best:Preschoolers use past reliability over age when learning new words. *Psychological Science, 17*(9), 757-758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology, 73*(1), 55-79. https://doi.org/10.1177/1747021819865833

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on

selective exposure to information. *Journal of Personality and Social Psychology, 80*(4), 557-571. https://doi.org/10.1037/0022-3514.80.4.557

Jones, C. S., & Kaplan, M. F. (2003). The effects of racially stereotypical crimes on juror decision-making and information-processing strategies. *Basic and Applied Social Psychology, 25*(1), 1-13. https://doi.org/10.1207/S15324834BASP2501_1

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28. https://doi.org/10.1037/0033-2909.114.1.3

Kassin, S. M., & Sommers, S. R. (1997). Inadmissible testimony, instructions to disregard, and the jury: Substantive versus procedural considerations. *Personality and Social Psychology Bulletin, 23*(10), 1046-1054. https://doi.org/10.1177/01461672972310005

Kaufmann, G., Drevland, G. C. B., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: displayed emotions and witness credibility. *Applied Cognitive Psychology, 17*(1), 21-34. https://doi.org/10.1002/acp.842

Lau, J. Y. F., Molyneaux, E., Telman, M. D., & Belli, S. (2011). The plasticity of adolescent cognitions: Data from a novel cognitive bias modification training task. *Child Psychiatry & Human Development, 42*(6), 679-693. https://doi.org/10.1007/s10578-011-0244-3

Loftus, E. F. (2003). Make-believe memories. *American Psychologist, 58*(11), 867-873. https://doi.org/10.1037/0003-066X.58.11.867

Lundrigan, S., Dhami, M. K., & Mueller-Johnson, K. (2018). A re-examination of the acquittal biasing effect of offence seriousness. *Psychiatry, Psychology and Law, 25*(5), 769-778. https://doi.org/10.1080/13218719.2018.1478334

Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*(5), 714-723. https://doi.org/10.1037/0021-9010.79.5.714

Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development, 79*(4), 914-929. https://doi.org/10.1111/j.1467-8624.2008.01167.x

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24*(15), 1315-1338. https://doi.org/10.1111/j.1559-1816.1994.tb01552.x

McCarthy Wilcox, A., & NicDaeid, N. (2018). Jurors' perceptions of forensic science expert witnesses: Experience, qualifications, testimony style and credibility. *Forensic Science International, 291*, 100-108. https://doi.org/10.1016/j.forsciint.2018.07.030

Mikolasch, S., Kotrschal, K., & Schloegl, C. (2011). African grey parrots (Psittacus erithacus) use inference by exclusion to find hidden food. *Biology letters, 7*(6), 875-877. https://doi.org/10.1098/rsbl.2011.0500

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Nitschke, F. T., McKimmie, B. M., & Vanman, E. J. (2019). A meta-analysis of the emotional victim effect for female adult rape complainants: Does complainant distress influence credibility? *Psychological Bulletin, 145*(10), 953-979. https://doi.org/10.1037/bul0000206

Pantell, R. H. (2017). The Child Witness in the Courtroom. *Pediatrics, 139*(3), 1-9. https://doi.org/10.1542/peds.2016-4008

Parrot may have the answer to a killing. (1993, November 12). *The New York Times*. https://www.nytimes.com/1993/11/12/news/parrot-may-have-the-answer-to-a-killing.html

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The Story Model decision and interference litigation. *Cardozo Law Review, 13*(Issues 2-3), 519-558. https://heinonline.org/HOL/P?h=hein.journals/cdozo13&i=543

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology, 62*(2), 189-206. https://doi.org/10.1037/0022-3514.62.2.189

Pepperberg, I. M. (1999). *The Alex studies: Cognitive and communicative abilities of grey parrots.* Harvard University Press. https://doi.org/10.2307/j.ctvk12qc1

Pepperberg, I. (2006). Grey parrot (Psittacus erithacus) numerical abilities: Addition and further experiments on a zero-like concept. *Journal of Comparative Psychology, 120*(1), 1-11. https://doi.org/10.1037/0735-7036.120.1.1

Pepperberg, I. M. (2007). Grey parrots do not always 'parrot': the roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Sciences*, 29(1), 1-13. https://doi.org/https://doi.org/10.1016/j.langsci.2005.12.002

Pica, E., Sheahan, C. L., Pozzulo, J., Vallano, J., & Pettalia, J. (2019). The influence of familiar and confident eyewitnesses on mock jurors' judgments. *Journal of Police and Criminal Psychology, 34*(4), 351-361. https://doi.org/10.1007/s11896-018-9306-9

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17*(1), 39-57. https://doi.org/10.1002/bdm.460

Rind, B., Jaeger, M., & Strohmetz, D. B. (1995). Effect of crime seriousness on simulated jurors' use of inadmissible evidence. *The Journal of Social Psychology, 135*(4), 417-424. https://doi.org/10.1080/00224545.1995.9712211

Rogers, P., Josey, N., & Davies, M. (2007). Victim age, attractiveness and, abuse history as factors in the perception of a hypothetical child sexual abuse case. *Journal of Sexual Aggression, 13*(2), 121-137. https://doi.org/10.1080/13552600701644835

Ross, D. F., Dunning, D., Toglia, M. P., & Ceci, S. J. (1990). The child in the eyes of the jury. *Law and Human Behavior, 14*(1), 5-23. https://doi.org/10.1007/BF01055786

Russo, J. E. (2014). The predecisional distortion of information. In *Neuroeconomics, judgment, and decision making* (pp. 109-128). Psychology Press. https://doi.org/10.4324/9781315763927

Ruva, C., McEvoy, C., & Bryant, J. B. (2007). Effects of pre-trial publicity and jury deliberation on juror bias and source memory errors. *Applied Cognitive Psychology, 21*(1), 45-67. https://doi.org/10.1002/acp.1254

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. *Journal of Applied Social Psychology, 34*(9), 1919-1944. https://doi.org/10.1111/j.1559-1816.2004.tb02593.x

Ruva, C. L., & Guenther, C. C. (2015). From the shadows into the light: How pretrial publicity and deliberation affect mock jurors' decisions, impressions, and memory. *Law and Human Behavior, 39*(3), 294-310. https://doi.org/10.1037/lhb0000117

Ruva, C. L., & Guenther, C. C. (2017). Keep your bias to yourself: How deliberating with differently biased others affects mock-jurors' guilt decisions, perceptions of the defendant, memories, and evidence interpretation. *Law and Human Behavior, 41*(5), 478-493. https://doi.org/10.1037/lhb0000256

Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes, 121*(2), 246-255. https://doi.org/https://doi.org/10.1016/j.obhdp.2013.02.001

Schloegl, C., Schmidt, J., Boeckle, M., Weiß, B. M., & Kotrschal, K. (2012). Grey parrots use inferential reasoning based on acoustic cues alone. *Proceedings of the Royal Society B: Biological Sciences*, 279(1745), 4135-4142. https://doi.org/doi:10.1098/rspb.2012.1292

Schuller, R. A., McKimmie, B. M., Masser, B. M., & Klippenstine, M. A. (2010). Judgments of sexual assault: The impact of complainant emotional demeanor, gender, and victim stereotypes. *New Criminal Law Review: An International and Interdisciplinary Journal, 13*(4), 759-780. https://doi.org/10.1525/nclr.2010.13.4.759

Schwarz, N. (2012). Feelings-as-information theory. In Fiske, S, T., Gilbert, D, T., & Lindzey, G (eds.), *Handbook of theories of social psychology*, *Vol. 1* (pp. 289-308). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n15

Siemaszko, C. (2016, December 3). Suspect in murder 'witnessed' by family parrot appears in court. *NBC News*. https://www.nbcnews.com/news/us-news/suspect-murder-witnessed-family-parrot-appears-court-n691271

Singh, R., Kumar, A., & Lehana, P. (2017). Investigations of the quality of speech imitated by Alexandrine parrot (Psittacula eupatria). *Circuits, Systems, and Signal Processing*, 36(6), 2292-2314. https://doi.org/10.1007/s00034-016-0395-3

Social Security Administration. (2019). *Top names of the 1960s*. https://www.ssa.gov/oact/babynames/decades/names1960s.html

Steblay, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior, 30*(4), 469-492. https://doi.org/10.1007/s10979-006-9039-7

Steblay, N. M., Besirevic, J., Fulero, S. M., & Jimenez-Lorente, B. (1999). The effects of pretrial publicity on juror verdicts: A meta-analytic review. *Law and Human Behavior, 23*(2), 219-235. https://doi.org/10.1023/A:1022325019080

Stöckl, H., Devries, K., Rotstein, A., Abrahams, N., Campbell, J., Watts, C., & Moreno, C. G. (2013). The global prevalence of intimate partner homicide: A systematic review. *The Lancet, 382*(9895), 859-865. https://doi.org/https://doi.org/10.1016/S0140-6736(13)61030-2

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755-769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tenney, E. R., Small, J. E., Kondrad, R. L., Jaswal, V. K., & Spellman, B. A. (2011). Accuracy, confidence, and calibration: how young children and adults assess credibility. *Developmental Psychology, 47*(4), 1065-1077. https://doi.org/10.1037/a0023273

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology, 44*(5), 1368-1375. https://doi.org/10.1016/j.jesp.2008.04.006

United States Census Bureau. (2016). *Technical Documentation: Demographic Aspects of Surnames - 2010 Census*. https://www2.census.gov/topics/genealogy/2010surnames/surnames.pdf

Voogt, A., & Klettke, B. (2017). The effect of gender on perceptions of credibility in child sexual assault cases: A systematic review. *Journal of Child Sexual Abuse, 26*(2), 195-212. https://doi.org/10.1080/10538712.2017.1280576

Voogt, A., Klettke, B., & Crossman, A. (2019). Measurement of victim credibility in child sexual assault cases: A systematic review. *Trauma, Violence, & Abuse, 20*(1), 51-66. https://doi.org/10.1177/1524838016683460

Voogt, A., Klettke, B., & Mohebbi, M. (2017). The development and validation of the child sexual assault victim credibility scale: An instrument to measure laypersons' perceptions of victim credibility. *Journal of Interpersonal Violence*. https://doi.org/10.1177/0886260517737554

Voogt, A., Klettke, B., & Thomson, D. (2017). The development of a conceptual model of perceived victim credibility in child sexual assault cases. *Psychiatry, Psychology and Law*. https://doi.org/10.1080/13218719.2017.1315764

Voogt, A., Klettke, B., Thomson, D. M., & Crossman, A. (2020). The impact of extralegal factors on perceived credibility of child victims of sexual assault. *Psychology, Crime & Law*, 1-26. https://doi.org/10.1080/1068316X.2020.1742336

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, 53(1), 5–13. https://doi.org/10.1037/0022-3514.53.1.5

Wiener, R. L., Arnot, L., Winter, R., & Redmond, B. (2006). Generic prejudice in the law: Sexual assault and homicide. *Basic and Applied Social Psychology, 28*(2), 145-155. https://doi.org/10.1207/s15324834basp2802_4

Willmott, D., Boduszek, D., Debowska, A., & Woodfield, R. (2018). Introduction and validation of the Juror Decision Scale (JDS): An empirical investigation of the Story Model. *Journal of Criminal Justice, 57*, 26-34. https://doi.org/10.1016/j.jcrimjus.2018.03.004

Wylie, B. E., Lyon, T. D., O'Connor, A. M., Aidy, C. L., & Evans, A. D. (2019). Adults' perceptions of children's referentially ambiguous responses. *Psychology, Crime & Law, 25*(7), 729-738. https://doi.org/10.1080/1068316X.2018.1552757

Zajac, R., Gross, J., & Hayne, H. (2003). Asked and answered: Questioning children in the courtroom. *Psychiatry, Psychology and Law, 10*(1), 199-209. https://doi.org/10.1375/pplt.2003.10.1.199

Zajac, R., & Hayne, H. (2003). I don't think that's what really happened: the effect of cross-examination on the accuracy of children's reports. *Journal of Experimental Psychology: Applied, 9*(3), 187-195. https://doi.org/10.1037/1076-898x.9.3.187

Zajac, R., & Hayne, H. (2006). The negative effect of cross-examination style questioning on children's accuracy: older children are not immune. *Applied Cognitive Psychology, 20*(1), 3-16. https://doi.org/10.1002/acp.1169

Zajac, R., O'Neill, S., & Hayne, H. (2012). Disorder in the courtroom? Child witnesses under cross-examination. *Developmental Review, 32*(3), 181-204. https://doi.org/10.1016/j.dr.2012.06.006

Appendix A

Vignette

Experiment 1 and 2

Two weeks ago, Mr. [husband name] [last name] contacted the police, saying that he had come home from work at 6 pm to find his wife in the backyard, covered in blood and dead.

An autopsy showed that Mrs. [last name] had been shot in the chest, and had quickly bled to death sometime between 1 pm and 3 pm. The police found no evidence at the scene to link the murder to any specific person.

Parrot Condition

Recently, investigators learned that the Wilsons' pet African Grey parrot, Leroy, had repeatedly been saying "**[husband name] did it**" since shortly after the murder occurred. Leroy was inside the house the day the crime occurred. Police now want to consider Leroy's statement in their investigation.

Child Condition

Recently, investigators learned that the Wilsons' 3-year-old step-son, Leroy, had repeatedly been saying "**[husband name] did it**" since shortly after the murder occurred. Leroy was inside the house the day the crime occurred. Police now want to consider Leroy's utterance in their investigation.

Experiment 3

Two weeks ago, Mr. [husband name] [last name] contacted the police saying that he had come home from work at 6 pm to find his wife in the backyard, covered in blood and dead.

An autopsy showed that Mrs. [last name] had been shot in the chest, and had quickly bled to death sometime between 1 pm and 3 pm. A neighbor reported that [husband name] and his wife had been fighting the night before. The police found no evidence at the scene to link the death to any specific person.

Highly ambiguous condition

Recently, investigators learned that [husband name]'s pet African grey parrot, Leroy, had repeatedly been saying "**[husband name], don't**" in the days following the death. Leroy was inside the house the day the death occurred. Police now want to consider Leroy's utterance in their investigation.

Less ambiguous condition

Recently, investigators learned that [husband name]'s pet African grey parrot, Leroy, had repeatedly been saying "**[husband name], don't shoot**" in the days following the death. Leroy was inside the house the day the death occurred. Police now want to consider Leroy's utterance in their investigation.

Appendix B

Original CSAVCS items compared to adapted scale items

Original CSAVCS (Voogt, Klettke, & Mohebbi, 2017)	**Adapted CSVCS**

Accuracy

1. This child's version of events was not accurate

2. This child will be able to give an accurate description of what happened to police

3. This child's memory for events was accurate

4. There was consistency within this child's version of events

5. There were inconsistencies in this child's report

Accuracy

1. Leroy's utterance is accurate

2. Leroy is able to give an accurate description of what happened to police

3. Leroy's memory for the event is accurate

—

—

Believability

1. This child's report was believable

2. This child was not believable

3. This child's report was credible

4. This child was not a credible witness

Believability

1. Leroy's utterance is believable

2. Leroy is not believable

3. Leroy's utterance is credible

4. Leroy is not a credible witness

Competency

1. This child was able to provide a competent account of events

2. This child is not a competent witness

3. This child had a good memory for the events

Competency

1. Leroy has the capacity to provide a competent account of events

2. Leroy is not a competent witness

3. Leroy has a good memory for the event

Reliability

1. This child is reliable

2. This child is dependable

3. This child's account of events can be depended upon

4. This child's memory of the event is likely to be unreliable

Reliability

1. Leroy is reliable

2. Leroy is dependable

3. Leroy's account of events can be depended upon

4. Leroy's memory of the event is likely to be unreliable

Truthfulness

1. This child was telling the truth about these events

2. This child's disclosure of what happened to her is truthful

3. This child was honest

4. This child's report was provided honestly

5. This child is trustworthy

6. I would trust a typical child of this age to be telling the truth

7. I believe the event occurred

Truthfulness

1. Leroy would tell the truth about this event if he had the language ability

2. Leroy's disclosure of what happened is truthful

3. Leroy is honest

4. Leroy's utterance is provided honestly

5. Leroy is trustworthy

6. I would trust a typical child of this age/ parrot of this species to be telling the truth

7. I believe the event occurred as Leroy said

Appendix C

Experiment 2 preregistration — https://aspredicted.org/blind.php?x=mp7xu9

CONFIDENTIAL - FOR PEER-REVIEW ONLY

ASPREDICTED

Non-Human Witness V3 Sept 2019 (#28133)

Created: 09/19/2019 02:26 PM (PT)
Shared: 07/19/2020 06:39 PM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review.
A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
To what extent do people believe a parrot is a credible witness, relative to a 3-year-old child?

3) Describe the key dependent variable(s) specifying how they will be measured.
Credibility: We will measure subjects' perceived credibility of the witness using an adapted version of the Child Sexual Assault Victim Credibility Scale (CSAVCS) (Voogt, Klettke & Mohebbi, 2017). Each item has been adapted to apply to both a parrot and child witness. We will use a 7-point Likert scale (1= Not at all, 7= Completely) to measure each item. We will calculate subscale scores as DVs.

Suspect's guilt: Subjects will be asked to rate the statement "Mr [husbands name] shot his wife" using a 6-point Likert scale (1 = Definitely not, 6 = Definitely).

Witness attempting to communicate: Subjects will be asked to rate the following statements on a 6-point Likert scale (1 = Definitely not, 6 = Definitely).
"[Witness] is attempting to communicate his own independent thoughts related to the shooting."
"[Witness] is attempting to communicate his own independent thoughts unrelated to the shooting."
"[Witness] is repeating something he heard related to the shooting."
"[Witness] is repeating something he heard that is unrelated to the shooting."

Relevance: Subjects will also be asked to rate the statement: "Leroy's utterance is relevant to the investigation" using a 7-point Likert Scale (1 = Not at all, 7 = Completely).

4) How many and which conditions will participants be assigned to?
Three between-subjects conditions: Subjects will be assigned to one of three witness conditions (Parrot, Child, or Control) randomly. Subjects will first read a vignette. The vignette describes a husband finding his wife shot and dead in the backyard. Additionally, the vignette states that the investigators currently have no leads or suspects. Subjects in the parrot and child condition will be informed that "Leroy" (either a parrot or child) is repeating a particular utterance (for example, "Paul did it "). Subjects in the control condition will not be informed about a witness.
We will counterbalance whether subjects complete the guilt rating or credibility ratings first. Subjects will rate what the witness is trying to communicate after the guilt and credibility ratings.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will conduct a between-subjects one-way ANOVA, and post-hoc tests, on subjects' rating of husband's guilt with witness type (child, parrot, control) as a between-subjects factor.
We will conduct a between-subjects Hotelling's two sample T2 test with witness type (child, parrot) as a between-subject factor and the credibility subscales as dependent variables
We will conduct a between-subjects independent t-test on each of the witness attempting to communicate items with witness type (child, parrot) as a between-subjects factor.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude from analyses the data from subjects who:
Do not complete the survey.
Fail the attention check.
Are inconsistent on the reversed questions.
Are incorrect about who Leroy is.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect data until we have complete, useable responses from 300 Mechanical Turk workers (~100 per condition).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

ASPREDICTED

We will run the credibility analysis with and without the truthfulness subscale.

As an attention check, we will ask the subjects who "Leroy" (the witness) is and will include an attention check question in the credibility scale. We will run all analyses both with and without excluding subjects who failed the attention check.

Available at https://aspredicted.org/blind.php?x=mp7xu9

Wharton UNIVERSITY of PENNSYLVANIA **CREDIBILITY LAB**

Experiment 3 preregistration — https://aspredicted.org/blind.php?x=di49bf

ASPREDICTED

Non-Human Witness V4 Feb 2020 (#34880)

Created: 02/02/2020 12:30 PM (PT)
Shared: 07/19/2020 06:41 PM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
When a parrot's utterance includes the name of the husband in a murder case, to what extent does the ambiguity of the utterance affect people's judgements about the husband's guilt?
We hypothesis that when the utterance is less ambiguous, people will be more likely to rate the witness as credible and the husband as guilty than when the utterance is more ambiguous or when there is no utterance.

3) Describe the key dependent variable(s) specifying how they will be measured.
Witness credibility: We will measure subjects' perceived credibility of the witness using an adapted version of the Child Sexual Assault Victim Credibility Scale (CSAVCS) (Voogt, Klettke & Mohebbi, 2017). We have adapted each item to apply to a parrot. Subjects will respond to each scale item using a 7-point Likert scale (1= Not at all, 7= Completely). We will use the mean of these 21 items as the witness credibility score.
Husband's guilt: Subjects will be asked to rate:
    Judicial guilt: "With regard to the crime of Mrs. [last name]'s murder, I think that Mr. [first name] [last name] is" (guilty/not guilty).
    Confidence in guilt rating: "How confident are you in your judgement that Mr. [first name] [last name] is [guilty/not guilty]" (1= not at all confident, 7= completely confident).
    Likelihood of guilt: "How likely is it that Mr. [first name] [last name] shot his wife? (1= very unlikely, 7= very likely)
Interpretation of utterance: Subjects will be asked to rate the following two statements on a 7-point Likert scale (1 = not at all, 7 = completely).
    "I believe the [witness]'s utterance is something he overheard."
    "I believe the [witness]'s utterance is related to the murder."

4) How many and which conditions will participants be assigned to?
Three between-subjects conditions: Subjects will be randomly assigned to one of three utterance conditions (high ambiguity, low ambiguity, no utterance). Subjects will first read a vignette. The vignette describes a husband who claims he found his wife shot and dead in the backyard. Additionally, the vignette states that the police found no evidence at the scene to link the death to any specific person. Subjects in the "low ambiguity" condition will be informed that "Leroy" is repeating the utterance "[husband's name], don't shoot". Subjects in the "high ambiguity" condition will be informed that "Leroy" is repeating the utterance "[husband's name], don't". Subjects in the "no utterance" control condition will not be informed about an utterance.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Following the estimation approach (Cumming, 2017) we will conduct a difference in proportions comparison between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' selection of the husband's judicial guilt (guilty/not guilty).
Following the estimation approach (Cumming, 2014) we will conduct planned comparisons between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' rating of the husband's likelihood of guilt.
We will conduct planned comparisons between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' ratings of confidence of guilt "How confident are you in your judgement that Mr. [first name] [last name] is [guilty/not guilty]".
We will conduct planned comparisons between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' ratings of witness credibility.
We will conduct planned comparisons between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' interpretation of utterance item "I believe the [witness]'s utterance is related to the murder."
We will conduct planned comparisons between each of the three conditions (high ambiguity, low ambiguity, no utterance) on subjects' interpretation of utterance item "I believe the [witness]'s utterance is something he overheard."

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude from analyses the data from subjects who:
Fail to complete the survey.
Fail either of the attention checks.
Are incorrect about who Leroy is.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Wharton UNIVERSITY of PENNSYLVANIA CREDIBILITY LAB

AS PREDICTED

We will collect data until we have 200 subjects per condition after exclusions from Mechanical Turk workers.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will collect basic demographics.
As an attention check, we will ask the subjects who "Leroy" (the witness) is, and will also include an attention check question in the credibility scale.

Wharton UNIVERSITY of PENNSYLVANIA CREDIBILITY LAB