

Do not fear your robot overlords: What a robot looks like does not affect how
much they mislead you

By
Gabriel Braniff

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Master of Science

Victoria University of Wellington

2015

Abstract

Would you believe what a robot tells you? A robot is essentially a high functioning computer and so, on the one hand, you should have no problem believing the information it provides. But on the other hand, what robots look like have been shown to affect how you would feel about them. Robots that look almost—but not quite—human have been shown to elicit feelings of unease and mistrust. How much do these feelings of mistrust in turn make these humanlike robots believed less? Across two experiments, we answered that question by showing people a video of a crime and then having them read a witness statement containing misleading information¹. This statement was ostensibly prepared by a robot that appeared human, robotic, or a morph of the two. Contrary to our predictions, what a robot looks like did not affect how misled people were, even when we drew attention to how much people trusted their robot source. These results suggest that even though people may not like, or trust, certain robots, they will still believe what they say.

¹ Although the research in this thesis is my own, I conducted it in a lab and supervised a team comprised of research assistants and honours students. I also received advice and direction from my supervisor. 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.

Acknowledgments

I would first of all like to thank my supervisor: Dr Maryanne Garry. The years I have spent working with you have been among the most challenging and rewarding of my life—I still maintain that you are scarier than any Sergeant Major in the military. There is no other person I would have rather worked with, and I appreciate every opportunity you gave me.

I would like to thank Robert, Mevagh, Brittany, Gregg, and the other members of the Garrylab. There is no other group of people that I could imagine spending years with, working on the same stuff day in and day out. You have been great friends as well as colleagues and my time at university would have been much less enjoyable without you all.

I would like to thank Anna for her friendship and support during Masters. Having someone to get a coffee with and have a break from research has been much appreciated.

Without you, or your inane(ly awesome) puns, Masters would have been dull.

I would also like to thank my parents, Judy and Mike; and my brothers, Phil and Jonny.

I really appreciate all your support and warm meals over the course of my studies.

Finally, I would like to thank the members of the Dragons & Lasers Floorball Franchise. Having a tight group to run out with each week has been invaluable for my sanity.

So thank you all, and #PPR.

Onward

~

Table of Contents

Abstract.....	i
Acknowledgments.....	ii
Table of Contents.....	iii
List of Tables and Figures.....	v
Introduction.....	1
Robot Appearance.....	1
The Misinformation Effect.....	3
Social Influence on the Misinformation Effect.....	4
Credibility.....	4
Social Power and Attractiveness.....	5
Overview of the Current Study.....	7
Norming of Materials.....	8
Method.....	8
Subjects.....	8
Design and Procedure.....	8
Results and Discussion.....	9
Experiment 1.....	11
Method.....	11
Subjects.....	11
Design.....	12
Procedure.....	12
Results and Discussion.....	15
Experiment 2.....	18

Method.....	18
Subjects.....	18
Design.....	18
Procedure.....	18
Results and Discussion.....	19
Meta-Analysis.....	23
General Discussion.....	24
References.....	28
Appendix A.....	32
Appendix B.....	33
Appendix C.....	36
Appendix D.....	37

List of Tables and Figures

<i>Figure 1</i> Examples of the three types of robots used by Prakash and Rogers (2013).....	3
<i>Figure 2</i> Mean ratings of trust for each of the four families of robots.....	9
<i>Table 1</i> Descriptive statistics for the norming study.....	10
<i>Figure 3</i> Diagram depicting the general method used in both Experiments.....	12
<i>Table 2</i> Example of four of the 24 counterbalances used in Experiment 1.....	13
<i>Figure 4</i> Mean accuracy and confidence on the memory test for Experiment 1.....	15
<i>Figure 5</i> Mean trust and accuracy for Experiment 2.....	20
<i>Figure 6</i> The misinformation effect for Experiment 2.....	22
<i>Figure 7</i> Meta-analysed misinformation effect.....	23
<i>Table A1</i> Images of the four families of robots used in the norming study.....	32
<i>Table A2</i> Counterbalancing for Experiment 1.....	36

Do not fear your robot overlords: What a robot looks like does not affect how much they mislead you

Would you believe what a robot tells you? This question might seem silly, but robots are becoming increasingly present in our lives. Whether they are vacuuming the floor of your apartment, or patrolling campuses in Silicon Valley, human-robot interaction is becoming more and more frequent. It is therefore not unrealistic to think that one day you might be approached by one of these campus-patrolling robots asking you questions about something you had witnessed. It is also not unrealistic for that robot to have malfunctioned—because of a hardware or software fault—and therefore to have recorded or analysed the event incorrectly and therefore have provided you with inaccurate information. Would you be misled by the information the robot provided? On the one hand, a robot designed to detect crime should have better vision than you, better hearing than you, and record things as they actually happened. So you might think of a robot as the ideal credible eyewitness and therefore believe what it says. But on the other hand, there is evidence to suggest that what a robot looks like—specifically how humanlike it looks—might influence how trustworthy you think it is.

Robot Appearance

The idea that a robot's appearance might affect how we feel about it is not new. In fact, since the 1970s, it has been theorised that how much we like a robot is dependent on how humanlike it appears (Mori, 1970; translated by McDorman & Minato, 2005; Mori, MacDorman, & Kageki, 2012). The more a robot resembles a human, the more familiar that robot becomes. But there is a point in which increasing how much a robot resembles a human no longer evokes a sense of familiarity, but rather one of eeriness or unease. Then, when that robot becomes indistinguishable from a human, that sense of unease fades and we treat it as we would any living person. That dip in familiarity has become known as *the Uncanny Valley*

(Mori, 1970). Since the original, unempirical, theory was proposed a number of studies have sought to empirically determine the extent to which how robots look affects what we think or how we feel about that robot.

Studies that investigate how a robot's appearance affects what we think about tend to follow a similar method (McDorman & Ishiguro, 2006; Prakash & Rogers, 2013; Burleigh, Schoenherr, & Lacroix, 2013). First, images of robots are digitally altered to vary in how much they resemble a human. Second, subjects rate these images on a number of characteristics, such as how familiar each appears, or how creepy each looks. In one such study, photos of androids—robots that look almost identical to a human—and robots were digitally morphed so that how much each morph resembled a human varied (McDorman & Ishiguro). When people were asked to rate how humanlike, familiar, or eerie each of these robots appeared, the more the morph resembled a human, the more humanlike and familiar people thought it appeared. But there was a point at which the more the robot resembled a human, the less familiar it became and instead felt eerie or creepy. Although this tipping point differed between studies, it does provide evidence that what a robot looks like affects how we feel about it. When considered as a whole, these data tell us that what we think or feel about that campus patrolling robot might be dependent on how much that robot resembles a human.

If what a robot looks like affects what we think or feel about it, then what a robot looks like might also affect how we interact with it. In one study that addressed this issue, photos of a human and robot were digitally combined to create a morph that resembled both the human and robot (Prakash & Rogers, 2013). Figure 1 shows an example of how these photos were altered. People were then asked which robot they would prefer to carry out various tasks such as at home providing social companionship, or helping them decide where to invest money. People preferred one of the two extremes of robots to help them with decision making, or providing social companionship. That is, they would prefer the robots that fell on the far left

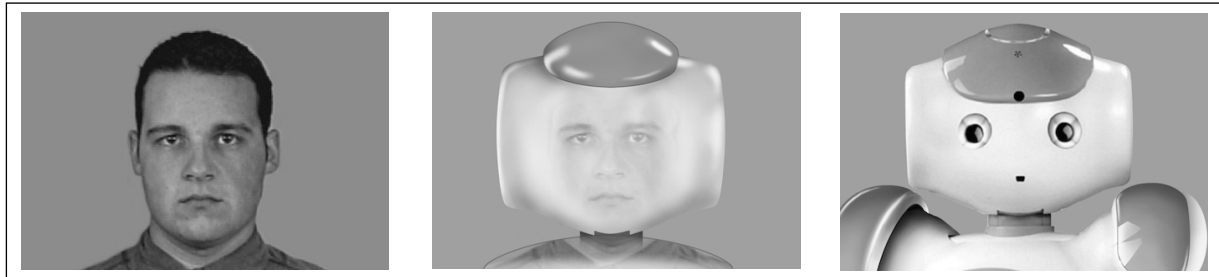


Figure 1. Examples of the three types of robots used by Prakash and Rogers (2013). The far left photo is a human, and the far right is an image of an actual robot. The middle image is a morph of both the human and robot. For the other images used, see Appendix A.

or far right of Figure 1 more than those robots that looked like the middle robot. And in fact, people also thought these middle robots were less trustworthy and likeable than the robots that were on either extreme.

Taken together, these studies suggest that not only does a robot's appearance affect how we feel about it, but it also affects how we would like to interact with that robot. And so we wondered how much that idea might also apply to robots that patrol campuses searching for crime. Would that robot's appearance also affect how much you would believe the information that it was telling you?

The Misinformation Effect

The large literature on eyewitness memory suggests that the answer is yes. Years of research have shown that receiving information after an event can alter what people report, even when the information provided contains misleading details (Loftus, Miller, & Burns, 1978; see Loftus 2005 for a review). The general method in the literature follows three phases. First, people are first shown an event, generally a slideshow or video depicting a crime. Second, after a delay, people are exposed to misleading information about what happened during that crime. Third, people are tested for their memory of the original event. Across hundreds of studies spanning decades of research, people tend to adopt that misleading *postevent information* and, when tested, report having seen it during the event. So

on the test, people are less accurate about those details they have been misled about (the misled items) than they are about details they have not been misled about (the control items). The extent to which these misleading details distort people's memory has become known as the *misinformation effect* (Tousignant, Hall, & Loftus, 1986).

The misinformation effect can be broken down into two distinct components: the cognitive component and the social component (Assefi & Garry, 2003). The cognitive component refers to how much people are able to remember about that specific event, outside of any misleading information. The social component, on the other hand, refers to how people's memory for the event changes in response to social influences (Assefi & Garry). The second phase of a typical misinformation experiment is almost always attributed to a person. Anything that influences how much people would believe that *misinformation messenger* also influences how much people believe the misleading information that messenger provides.

Social influence on the misinformation effect

Credibility. One of the factors that influences the extent to which people succumb to misleading postevent information is how credible the misinformation messenger is. The more credible the misinformation messenger, the more likely people are to believe the information that the messenger provides (Dodd & Bradshaw, 1980; French, Garry, & Mori, 2011; Zajac, Dickson, Munn, & O'Neill, 2013). In one study, subjects read a postevent narrative from the point of view of either an innocent bystander who witnessed the crime, or the suspect's defence attorney (Dodd & Bradshaw). People who read the defence attorney's narrative were less likely to falsely report the misleading information contained in the narrative compared to those who read the innocent bystander's report.

Perhaps it is not surprising that people would be less likely to believe the information a defence attorney provides, because defence attorneys are not seen as a credible source of information about a crime. But there is also evidence to suggest that how credible you think

you are relative to the person providing you with postevent information affects how likely you are to be misled (French et al., 2011; Zajac et al., 2014). To address this idea, one study showed subjects a video depicting a crime, and then had them discuss the crime with someone who had also watched the video with them (French et al., 2011). Before watching the video, subjects were each given a set of glasses to wear and were told that these glasses would reduce their visual acuity. Subjects were then led to believe that their visual acuity was better, worse, or equal to the person who was also watching the video with them. In order to ensure subjects saw different versions of the film and could therefore mislead each other, the Mori technique was used where two versions of a film were displayed on the same screen (Mori, 2007). The glasses that subjects were told would reduce their visual acuity were in fact used to filter out one version of the film, thereby ensuring each subject unknowingly saw a different version of the video. After the video, subjects were asked questions that prompted a discussion about the scenes in the video that were different between versions. When their memory for the video was later tested, subjects who thought their eyesight was worse than the other witness were more likely to believe what the other witness was saying was correct and therefore more likely to report these misleading details in a memory test. By contrast, subjects who thought their eyesight was better than the other witness were less likely to be misled by that witness, and therefore tended to be more accurate on the test. These data suggest that when people think they are a more credible eyewitness, they are less likely to believe the information a less credible eyewitness provides them.

Social power and attractiveness. We probably think defence attorneys are motivated to portray their clients as innocent, and we expect that people who had a better view of the crime would have information than others who had a relatively poorer view. But a seemingly irrelevant piece of social information also influences the misinformation effect: someone's accent. Although an accent provides no information about how reliable or credible a source

is, it does provide a number of social cues, such as power and attractiveness (Frumkin, 2007; Fuertres, Gottdiener, Martin, Gilbert, & Giles, 2011). In one study, subjects watched a slideshow depicting a mock crime, and then listened to postevent information delivered in either a New Zealand, or North American accent (Vornik, Sharman, & Garry, 2010). When people thought the misinformation messenger conveyed power, it did not matter how socially attractive the messenger was: subjects were all similarly misled. But when subjects thought the messenger conveyed little power, then those who found the messenger socially attractive were more misled than those who thought the messenger was relatively less socially attractive. And so even though a person's accent does not provide any true diagnostic information about the credibility of a source, the social power and attractiveness that people associate with a given accent does in fact influence how credible people think a source is.

Together these studies provide some support for the idea that characteristics of robots—such as what they look like—might affect how likely they are to mislead us. In fact, one study addressed this issue (Bethel, Eakin, Anreddy, Stuart, & Carruth, 2013). In the only other study to date, people first watched a slideshow depicting a crime, then were interviewed by either a human, or robot. That interview consisted of leading questions designed to mislead the subject about certain details in the crime. Overall, people who were interviewed by a human showed the usual misinformation effect, but people who were interviewed by the robot showed no misinformation effect. At first glance, this study suggests that perhaps robots are not able to deceive us. But there is a problem with the study. People interacted with the robot both before and after the interview. Allowing people to play with a robot may have reduced the credibility of the robot as an eyewitness because it reduced the robot to a simple toy. And so, when the robot attempted to mislead them, people were more likely to find the inconsistencies in the information it provided and therefore were less prone to being misled.

Or alternatively, they simply did not take the robot's interview seriously—if they viewed the robot as a toy—and so were not affected by the misleading questions.

Taken together, these studies suggest that a number of characteristics about a source of information can affect how susceptible we are to being misled by that source. Even when these characteristics are seemingly irrelevant, the social information they convey can be enough to affect a source's credibility.

Overview of current study

Considered as a whole, these two lines of research suggest that what campus-patrolling robots look like might affect how much we trust them. How much we trust those robots might in turn affect how likely we are to believe the information they provide us. And so we wondered how much a robot that had been digitally altered to resemble a human would be less likely to mislead people than robots that were identical to a human, or resembled what people think a robot should look like. Put another way: to what extent does what a robot looks like affect how much it can mislead you? That is the primary research question we address in this thesis.

To answer that question, we first showed subjects a video of a mock crime, and then had them read a witness statement that was ostensibly prepared by one of three types of robots; a robot that looked human, a robot that looked robotic, or a robot that was a digital morph of a human and robot (see Figure 1 for an example). Contained in this witness statement were eight critical sentences which either contained misleading information, or no specific details at all. After reading the witness statement, people's memory for the original event was tested. If what a robot looks like affects how much people are misled by it, then when people receive misleading postevent information from a robot they trust, they should be more likely to succumb to those misleading details and be less accurate on a subsequent

memory test compared to those who read postevent information from a robot they do not trust.

Before investigating the extent to which a robot's appearance affects how misled people are by them, we first wanted to know if the morphs being used in the study were trusted less than their human, or robot counterparts. We also wanted to identify two “families,” or groups of human, robot, and morph, that showed the greatest difference in trust between these three robots. We therefore ran a norming study where we simply asked people to rate how much they trusted each robot.

Norming of materials

Method

Subjects. We recruited 309 subjects using Amazon’s Mechanical Turk (MTurk; www.mturk.com/mturk) service. Subjects received 0.5 USD in Amazon gift vouchers in exchange for their participation. Studies run online using MTurk attract diverse subjects and produce similar results to studies run in the laboratory (Buhrmester, Kwang, & Gosling, 2011; Germine et al., 2012; Mason & Suri, 2012). We originally intended to collect 300, but due to a quirk in the way our data collection software (www.Qualtrics.com) interacts with MTurk that led to slight over-collection, we actually collected data from 309 people (Age: 18 - 73, $M = 33.58$, $SD = 12.11$).

Design and Procedure. We sourced images of humans and robots that had been digitally combined to create a human-robot morph (Prakash & Rogers, 2013). We grouped these human-robot Morphs into “families” of three robots; the Morph, and both the Human and Robot that the morph was created from. We randomly presented subjects one image from each of the four families (see Appendix A for images of all robots).

We then asked subjects: a) how useful would you find a robot with this appearance? b) how much would you trust a robot with this appearance? c) how much would you like a robot

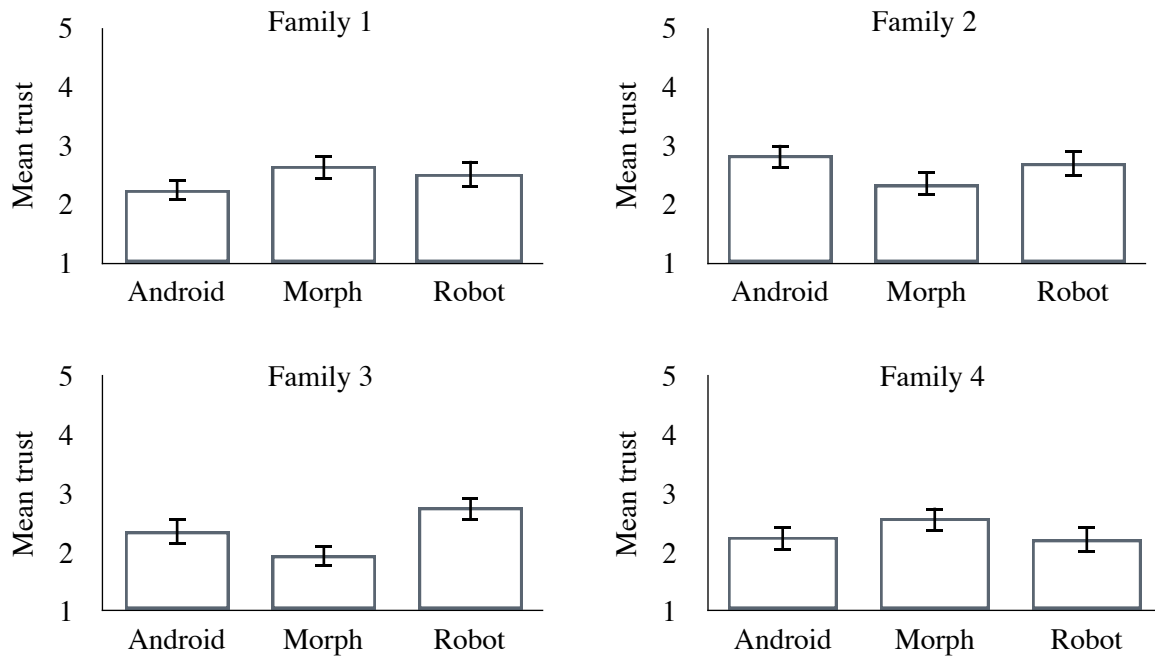


Figure 2. Mean ratings of trust for each of the four families of robots. Error bars represent 95% CIs for each cell mean.

with this appearance? d) how anxious would a robot with this appearance make you? Subjects responded on a 1 (*Not at all*) to 5 (*Very much*) scale. Both the order and type of robot that subjects saw was completely randomised.

Results and Discussion

Our primary question of interest was: to what extent does what a robot look like affect what people think about that robot? But before we address this question we first identified subjects who failed to complete the experiment, $n = 8$, (3%) and excluded them from all analyses. This process left us with 301 subjects in the dataset.

We next calculated subjects' mean responses for the four questions and grouped these responses according to which family each robot belonged to and the type of robot it was (Android, Morph, or Robot), we display these descriptive statistics in Table 1, and plot subjects mean trust scores in Figure 2.

In order to find two families where the Morph was trusted less than both the Android and Robot—a 'V' shape pattern—we next took subjects responses to the question about how

Table 1

Descriptive statistics for the norming study

Robot	Trust		Like		Usefulness		Anxiety	
	Mean (<i>SD</i>)	95% CI	Mean (<i>SD</i>)	95% CI	Mean (<i>SD</i>)	95% CI	Mean (<i>SD</i>)	95% CI
Family 1								
Android	2.24 (1.05)	[2.03, 2.44]	2.17 (1.13)	[1.94, 2.39]	2.67 (1.19)	[2.45, 2.92]	2.57 (1.34)	[2.31, 2.84]
Morph	2.63 (1.11)	[2.41, 2.84]	2.45 (1.22)	[2.21, 2.69]	2.91 (1.07)	[2.70, 3.12]	2.75 (1.25)	[2.50, 3.00]
Robot	2.52 (1.16)	[2.29, 2.75]	2.61 (1.24)	[2.36, 2.86]	2.68 (1.17)	[2.45, 2.91]	2.59 (1.32)	[2.33, 2.85]
Family 2								
Android	2.82 (1.23)	[2.58, 3.06]	2.86 (1.28)	[2.61, 3.11]	3.18 (1.26)	[2.94, 3.43]	2.84 (1.42)	[2.57, 3.12]
Morph	2.34 (1.08)	[2.12, 2.55]	2.29 (1.12)	[2.06, 2.51]	2.40 (1.12)	[2.17, 2.62]	2.78 (1.43)	[2.49, 3.06]
Robot	2.70 (1.15)	[2.47, 2.93]	2.81 (1.22)	[2.57, 3.05]	2.68 (1.12)	[2.46, 2.90]	2.14 (1.06)	[1.93, 2.35]
Family 3								
Android	2.34 (1.18)	[2.10, 2.57]	2.33 (1.22)	[2.08, 2.57]	2.77 (1.33)	[2.50, 3.03]	2.92 (1.40)	[2.64, 3.20]
Morph	1.95 (0.94)	[1.77, 2.14]	1.68 (0.91)	[1.50, 1.86]	1.94 (0.96)	[1.75, 2.13]	2.83 (1.30)	[2.58, 3.09]
Robot	2.75 (1.02)	[2.55, 2.96]	2.90 (1.23)	[2.66, 3.14]	2.60 (1.05)	[2.40, 2.80]	1.95 (1.04)	[1.75, 2.15]
Family 4								
Android	2.26 (1.06)	[2.05, 2.47]	2.22 (1.14)	[2.00, 2.45]	2.69 (1.26)	[2.44, 2.94]	2.82 (1.34)	[2.55, 3.09]
Morph	2.56 (1.08)	[2.35, 2.77]	2.40 (1.08)	[2.19, 2.61]	2.83 (1.06)	[2.62, 3.04]	2.59 (1.15)	[2.36, 2.82]
Robot	2.22 (1.01)	[2.03, 2.42]	2.07 (1.10)	[1.85, 2.28]	2.29 (1.00)	[2.10, 2.49]	2.84 (1.27)	[2.60, 3.09]

Note: 95% CIs refer to the 95% confidence intervals for each cell mean.

much they would trust each robot and calculated a mean trust score and display these data in

Figure 2. As Figure 2 shows, only two families, two and three, showed a pattern of results

consistent with what we would expect, in which the Morphs were trusted less than the

Android or Robot. More specifically, for Family 2, the Morph was trusted less than both the

Android, $M_{\text{diff}} = 0.48$, 95% CI [0.09, 0.86], and Robot, $M_{\text{diff}} = 0.36$, 95% CI [-0.02, 0.75]. The

Android and Robot were similarly trusted, $M_{\text{diff}} = 0.12$, 95% CI [-0.27, 0.50]. For Family 3

the pattern was slightly different; consistent with Family 2, the Morph was trusted less than

both the Android, $M_{\text{diff}} = 0.39$, 95% CI [0.04, 0.74], and Robot, $M_{\text{diff}} = 0.80$, 95% CI [0.46,

1.15], but this time the Android was also trusted less than the Robot, $M_{\text{diff}} = 0.42$, 95% CI

[0.07, 0.77]. The critical finding in this study, however, is that the Morph was the robot trusted the least.

In Null Hypothesis Significance Testing (NHST) terms, one way ANOVAs revealed a main effect of Type of Robot for Family 2, $F(2, 300) = 4.65, p = .01$, and Family 3, $F(2, 301) = 14.96, p < .001$. Even though these ANOVAs provided a main effect for Type of Robot for Family 1, $F(2, 303) = 3.43, p = .03$, and Family 4, $F(2, 302) = 3.13, p = .05$, the pattern of results were in the opposite direction to what we were looking for: the Morphs were actually trusted more than the Android or Robot.

In fact, when we consider the other four questions in the study: how useful people thought each robot was; how much people liked each robot; and how anxious each robot made people, we get a consistent pattern of results: for families two and three, people liked the Morph less and thought it was less useful than the Android or Robot. People also thought the Morph was more anxiety provoking than either the Android or Robot.

Now that we have two families of robots that show a pattern of trust consistent with what we expected, we next investigated the extent to which this relatively less trust would translate into people being less misled by these robots. We address this issue in the two experiments that follow.

Experiment 1

Method

Subjects. Subjects were recruited through Amazon's MTurk and received 1 USD for their participation. Based on pilot testing, we predetermined a sample size of 300 but due to a programming error when coding the experiment, 136 people were unable to complete the

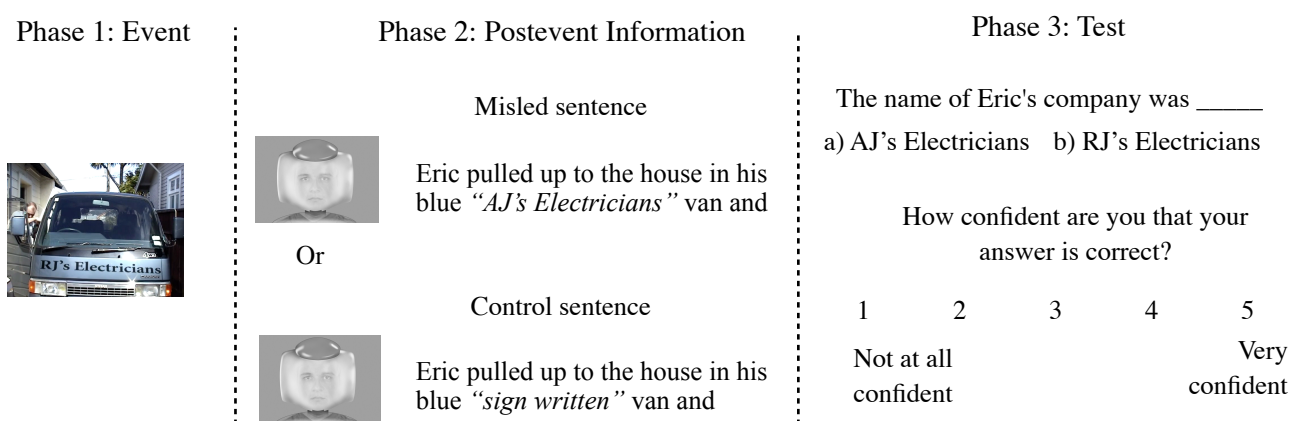


Figure 3. Diagram depicting the general method used in both Experiments. In the Postevent information phase, the top sentence is an example of a misleading sentence and the bottom sentence is an example of a control sentence with italics illustrating the difference between the two types of sentences.

experiment. We therefore corrected that error and collected data from a further 335 people (Age: 18 - 69 ($M = 34.12$, $SD = 11.53$)²).

Design. We used a 3 (Source: Android, Morph, Robot) x 2 (Item Type: Control, Misled) mixed factors design with Source manipulated between-subjects.









Procedure. We adapted materials from Takarangi, Parker, and Garry (2006) and as Figure 3 shows, the experiment proceeded in three phases.

Phase 1. Subjects watched a 6min and 36s silent film of an electrician (Eric) stealing items and rummaging through the personal belongings of a client's unattended house (Takarangi et al., 2006). Subjects saw one of two versions of the film, identical except for eight digitally altered critical scenes. For example, in one version of the film subjects saw Eric pull up to the house in a blue van emblazoned with "RJ's Electricians". But in the other version of the film, subjects saw Eric pull up to the house in the same van but with the words "AJs Electricians" emblazoned on the front. The video version was counterbalanced so that equal numbers of subjects saw either version.

² We aimed to collect data from 300 people but due to the same quirk as the pilot study, we actually collected data from 335 people.

Table 2

Example of four of the 24 counterbalances used in Experiment 1

Counterbalance	Video Version	Phase 2	
		Robot	Narrative example
1			Eric pulled up to the house in his blue “ <i>AJ's Electricians</i> ” van and parked in the driveway.
2			Eric pulled up to the house in his blue <i>sign written</i> van and parked in the driveway.
3			Eric pulled up to the house in his blue “ <i>AJ's Electricians</i> ” van and parked in the driveway.
4			Eric pulled up to the house in his blue <i>sign written</i> van and parked in the driveway.

Note: This pattern was repeated for each of the robots in both families. In Experiment 2 the counterbalance was identical but for only one family of robots. For a full counterbalancing table, see Appendix C.

When subjects had finished watching the video, they played an online card matching game for 10 minutes (www.coghub.com/memory/memory.html). The card matching game involved flipping pairs of cards until subjects found two that matched. Subjects continued playing until the 10 minutes had passed—the game reset if they matched all the pairs within 10 minutes. This game served as a filler task in order to allow memory for the video to decay.

Phase 2. After the filler task, we told subjects that a robot had also witnessed the crime they had just seen in the video and that they were about to read this robot’s witness statement. Included in the witness statement were eight sentences referring to the eight critical scenes in the video. Four of these sentences were control sentences, describing a critical scene of the video generically, without any descriptive information (e.g. Eric pulled up to the house in his blue sign written van and parked in the driveway). The remaining four critical sentences were

misleading sentences, described a scene specifically but with misleading information (e.g. Eric pulled up to the house in his blue “AJ’s Electricians” van and parked in the driveway). These sentences were counterbalanced such that each sentence appeared equally often as either a control or misleading sentence.

The witness statement appeared one sentence at a time alongside a photo of the robot who was providing the witness statement (see Table 2). Subjects could read each sentence of the witness statement in their own time but could not return to earlier sentences once they had moved on. The robots used in this study were the two families of robots we identified in our norming study that showed the greatest difference in trust for the Morph relative to the Android and Robot. After reading the witness statement, subjects played the same card matching filler task but this time for only 3 minutes.

Phase 3. After the second filler task, subjects were given a surprise memory test for the video itself. The test consisted of 20 two-alternative forced choice questions (e.g. The name of Eric’s company was ____) where subjects could choose between the correct or misleading answer (AJ’s Electricians or RJ’s Electricians). After each question we also asked "How confident are you that your answer is correct." Subjects responded on a 1 (*Not at all confident*) to 5 (*Very confident*) scale. The test included eight critical questions referring to the four control and four misleading sentences in the witness statement, the remaining 12 questions served as filler items.

After the memory test, we asked subjects if they had seen the video of Eric the electrician before, or if they had taken notes at any stage during the experiment. To encourage subjects to respond honestly, we told them we would pay them regardless of their answers to these question. Subjects were then debriefed in full.

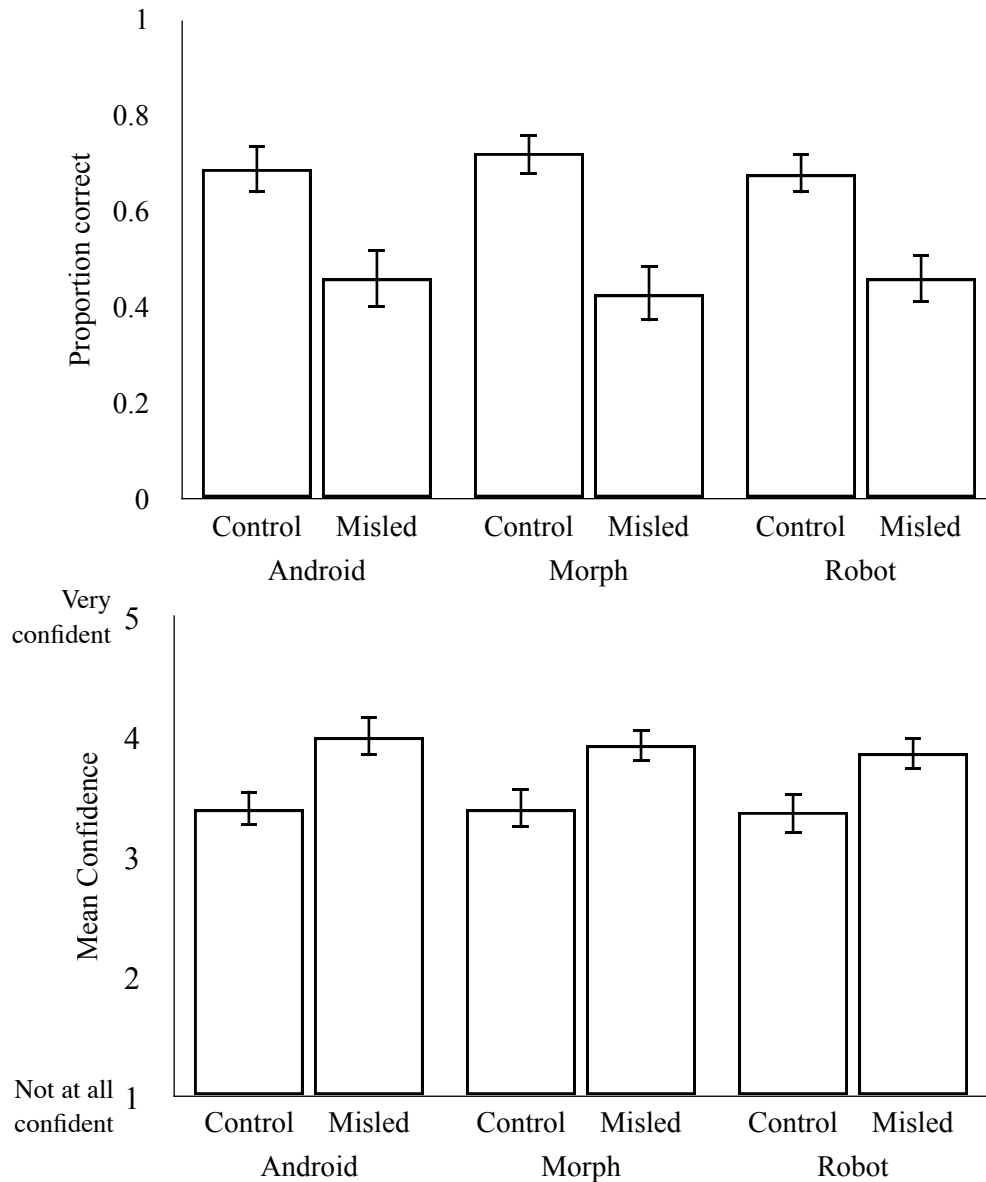


Figure 4. Mean proportion correct (Accuracy) on the memory test (Top panel) and mean confidence (bottom panel) as a function of Item Type and Source in Experiment 1. Error bars represent 95% CIs.

Results and Discussion

Before we address our primary question, we first identified those subjects who had seen the video of Eric the electrician before, $n = 10$ (3%), or who had taken notes at any stage during the Experiment, $n = 8$ (2%), and exclude these subjects from all analyses. This procedure left us with 317 people in our dataset.

We now turn to our primary question: to what extent did what a robot looks like affect how much it misled subjects? To answer that question, we calculated subjects' mean accuracy

and confidence ratings for the eight critical questions and classified them first according to whether they referred to control or misleading details, and second according to the type of robot eyewitness that provided them the witness statement. We display subject's mean accuracy (top panel) and mean confidence (bottom panel) in Figure 4.

The top panel of Figure 4 illustrates two important findings. First, as is shown when we look at performance on the control items, when subjects were not exposed to any misleading information they did reasonably well on the memory test. This performance on the control items suggests that their memory for the events in the video was relatively good. But when subjects were exposed to misleading information, their performance on the memory test was reduced: they were more likely to respond with the misleading details rather than the accurate details. Second, and contrary to what we predicted, what a robot looked like had no meaningful influence on people's tendency to be misled by it. That is, regardless of what the robot looked like, it misled people to a similar degree.

In other words, for the control items, subjects who read the Morph's statement performed similarly well on the memory test compared to those who read the Android's, $M_{\text{diff}} = 0.03$, 95% CI [-0.04, 0.10], and Robot's, $M_{\text{diff}} = 0.04$, 95% CI [-0.03, 0.11]. This pattern of results was the same for the Misled items, subjects who read the Morph's statement performed equally well on the memory test compared to those who read the Android's, $M_{\text{diff}} = -0.02$, 95% CI [-0.07, 0.12], or Robot's, $M_{\text{diff}} = 0.03$, 95% CI [-0.06, 0.12]. In NHST terms, a mixed-factors ANOVA revealed only the misinformation effect, which is shown by the main effect of Item Type, $F(1, 314) = 3499.97$, $p < .001$.

Did this tendency to be misled by the eyewitness affect how confident people were in their answers? The answer is yes. As the bottom panel of Figure 4 shows, overall people were more confident their answers to the questions they had been misled about were correct than the control questions. But, as was the case with accuracy, the type of robot that provided the

witness statement only trivially affected how misled people were. Specifically, people who read the Morph's witness statement were similarly confident in their answers to the Control questions than those who read the Android's, $M_{\text{diff}} = 0.01$, 95% CI [-0.27, 0.28], or Robot's, $M_{\text{diff}} = 0.03$, 95% CI [-0.24, 0.31]. People who read the Morph's statement were also similarly confident in their answers to the Misled questions as those who read the Android's, $M_{\text{diff}} = 0.04$, 95% CI [-0.18, 0.27], or Robot's, $M_{\text{diff}} = 0.06$, 95% CI [-0.17, 0.29] Or, in NHST terms, a mixed-factors ANOVA revealed only a main effect for Question Type, $F(1, 314) = 132.69$, $p < .01$.

These results suggest that even though we do not trust robots that fall into the Uncanny Valley as much as those that do not, this lack of trust does not translate into a difference in susceptibility to the misinformation effect. It appears as though the appearance of a robot provides little to no diagnostic information about how much we should believe what it tells us. But why do these robots that we trust less not also mislead us less? One possible explanation is that there needs to be a comparison between robots before these differences in trust occur. In both our norming study and the study we sourced our materials from, people were able to see all the robots used in the study and therefore directly compare them (Prakash & Rogers, 2013). And so the reason why the Morphs are trusted less might be because people felt they would trust those robots less than either the Android or Robots they had just seen. But when we took only the ratings of trust for the first robots people saw in the norming study, the pattern of results did not change: people still trusted the Morphs in Families 2 and 3 less than the Android or Robot. To investigate this possibility, in Experiment 2 we show subjects all the photos for the family of robots and told them one of these robots would be providing them a witness statement.

An alternative explanation is that subjects were not explicitly aware of how much they trusted the robot during the experiment. We know that there are a number of factors that

affect how people rely on feelings—such as trust—as a source of information (see Greifeneder, Bless, & Pham, 2011, for a review). Included in these factors are how salient the feeling is, and how relevant that feeling is to the task. In order for people to use feelings as information, that feeling has to stand out enough that they are aware of it, and it has to be relevant to the judgment they are making, otherwise there is little reason to use that feeling. One way that we can increase the salience and relevance of how much people trust our robots and encourage them to use that information, is to ask them about how much they trust the robot prior to reading its witness statement (Kuhnen, 2010). We also address this alternative explanation in Experiment 2.

Experiment 2

Method

Subjects. We recruited 115 people through Victoria University of Wellington's Introduction to Psychology Research Programme (IPRP). IPRP subjects are students enrolled in an introductory psychology course who participate in psychological experiments in exchange for course credit. Subjects received 0.5 credits towards this mandatory course requirement. We initially aimed to recruit 150 subjects but due to insufficient volunteers, we were only able to collect data from 115 people.

We also recruited 192 people through Amazon's MTurk. MTurk workers received 0.5 USD in exchange for their participation. We aimed to collect data from 150 people, but due to the same quirk as in the norming study, actually collected data from 192 people.

The data from both sources were combined to create a single dataset, $N = 307$ (Age: 17 - 76, $M = 29.82$, $SD = 13.39$).

Design. We used a 3 (Source: Android, Morph, Robot) x 2 (Item Type: Control, Misled) mixed factors design with Source manipulated between-subjects.

Procedure. Experiment 2 was identical to Experiment 1 with the following exceptions:

To increase the power of our study, we collected data only for one Family of robots. We chose Family 3 because norming studies revealed the largest difference between the Morphs and the Android/Robot counterparts. In addition, to allow subjects to compare the three robots in the family, we showed subjects all three robots and told them “Because the event took place at the home of a robotics engineer, the above robots also witnessed what you just saw in the video.” Subjects read the witness statement of only one of these three potential sources.

Finally, to draw attention to how much subjects trusted their source of information, we asked subjects how much they trusted the robot whose witness statement they were about to read. Subjects responded on a 1 (*Not at all*) to 5 (*Very much*) scale.

Results and Discussion

We first identified subjects who failed to complete the experiment, $n = 43$ (14%), had seen the video of Eric the electrician before, $n = 6$ (2%), took notes at any stage during the experiment, $n = 2$ (0.7%), or who revealed they had rewind the video to re-watch it, $n = 6$ (2%), and exclude them from all analyses. This process left us with 250 people in the dataset.

We next wanted to know if our manipulation worked, or how much what a robot looks like affected how much people trust them. To answer that question, we took subjects’ responses to the question about how much they trusted the robot providing them a witness statement and grouped them according to the robot eyewitness. We display these data in the top panel of Figure 5. And as the figure shows, the manipulation worked. People trusted the Morph less than both the Android, $M_{\text{diff}} = 0.27$, 95% CI [-0.08, 0.63], and Robot, $M_{\text{diff}} = 0.39$, 95% CI [0.03, 0.75]. But these differences were small, differing by only a quarter of a point on a 5-point scale, meaning each robot differed by only about 6%. In fact, when comparing how much people trusted the Android and Morph, the confidence intervals surrounding that difference cross zero, suggesting that plausibly there is no difference at all, or even one in the

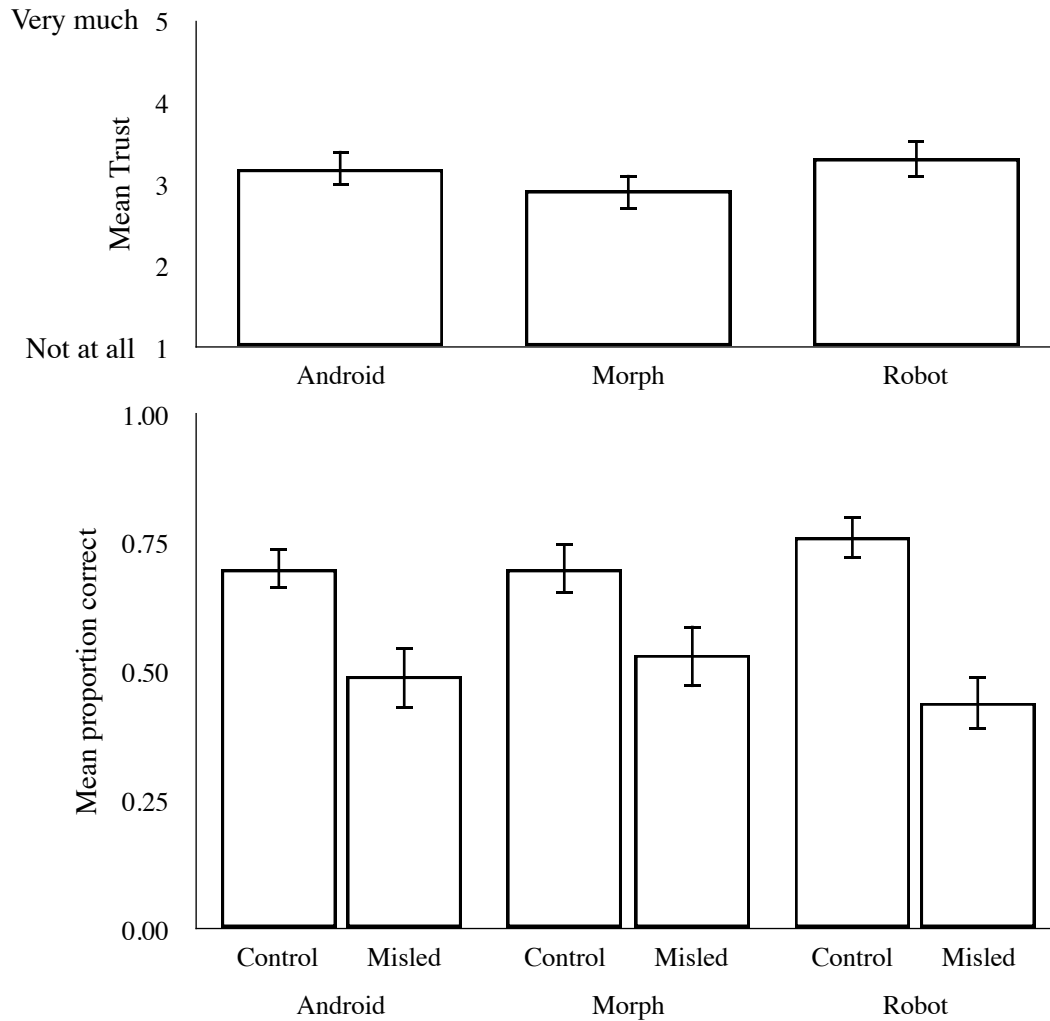


Figure 5. Mean ratings of trust (top panel) and mean accuracy (bottom panel) as a function of Type of Robot and Item Type (Accuracy only). Error bars refer to 95% CIs.

opposite direction. In NHST terms, a one-way ANOVA revealed a main effect of Source, $F(2, 247) = 3.43, p = .03$.

And now our primary question: to what extent does what a robot looks like affect how likely people are to believe information it provides when they are explicitly aware of how much they trust that robot? To answer this question, we calculated subjects' mean accuracy scores for the eight critical questions and classified them in the same way as Experiment 1. We display these data in the bottom panel of Figure 5. The answer, as Figure 5 shows, is that it depends. Like Experiment 1, the bottom panel of Figure 5 tells us a number of important findings: First, when people were not exposed to misleading information, everyone did

reasonably well. Second, when subjects were exposed to misleading information, their performance on the memory test decreased, suggesting that people had again been misled by the robot eyewitness. Third, unlike in Experiment 1, what the robot eyewitness looks like affected people's performance on the test. Specifically, people who read the Robot's statement now did trivially better on the Control items, Android: $M_{\text{diff}} = 0.06$, 95% CI [-0.02, 0.14]; Morph: $M_{\text{diff}} = 0.06$, 95% CI [-0.02, 0.14], and slightly worse on the Misled items, Android: $M_{\text{diff}} = 0.06$, 95% CI [-0.05, 0.16]; Morph: $M_{\text{diff}} = 0.09$, 95% CI [-0.01, 0.19], than subjects who read the Android's or Morph's. The confidence intervals spanning these differences, however, cross zero, suggesting that there is no plausible difference between these groups. Although there might be no plausible differences, the pattern of results we get fits with the pattern of results we obtained for people's ratings of trust: people trusted the robot the most and were also the most misled by it.

In other words, a mixed-factor ANOVA revealed a Source by Item Type interaction, $F(2, 247) = 4.49, p = .01$. This interaction should be interpreted cautiously, however, as there were no significant differences between the robot eyewitnesses on either the Control or Misled items. Instead, what is driving the interaction found in this experiment is the size of the difference between those Control items and Misled items in each group. To illustrate this explanation, we calculated the difference between the proportion correct for Control and Misled items and plot these differences in Figure 6. To interpret Figure 6, the y-axis can be conceptualised as the size of the misinformation effect—the higher the bars, the greater the difference between control and misled items, and hence the stronger the misinformation effect. As the figure shows, when we control for people's performance on the Control items, subjects who had read the Robot's witness statement were the most susceptible to the misinformation effect. Whereas subjects who read the Morph or Android's statement were similarly susceptible to the misinformation effect.

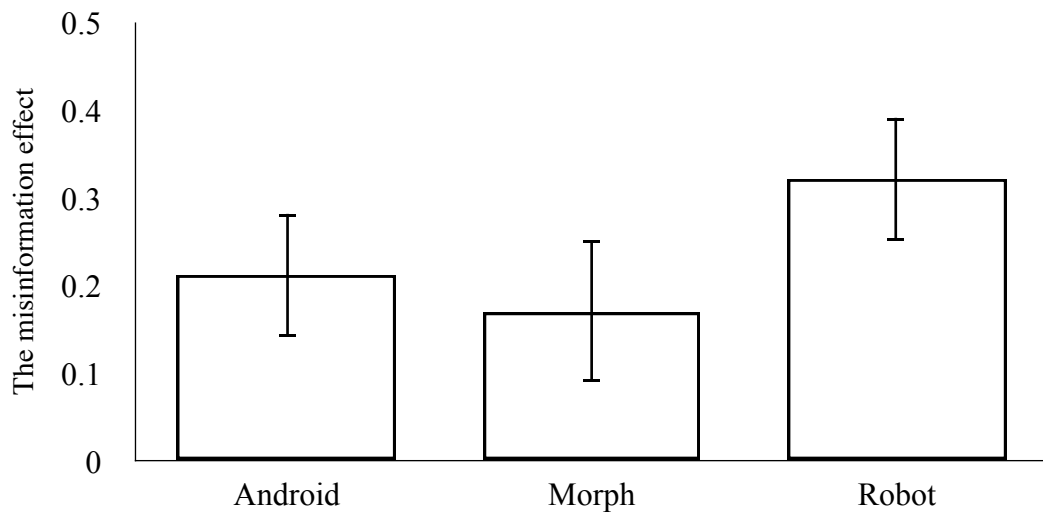


Figure 6. The misinformation effect. The difference of proportion of correct scores between Control and Mislead items as a function of Type of Robot. Error bars represent 95% CIs.

These data suggest that when we drew attention to how much subjects trust the robot, and provided them with a comparison, then the appearance of the robot whose witness statement subjects were reading then affected how misled they were. But the pattern of results we get with people's accuracy is not consistent with what we would expect given their trust ratings. People trusted the Morph less than the Android and Robot, and so should have the smallest misinformation effect. Instead, the size of the misinformation effect is similar for both the Android's and Morph's, with the Robot having the largest misinformation effect.

But another test for how much trust affects the misinformation effect is to investigate the extent to which people's ratings of trust are related to the size of the misinformation effect. If, as expected given the literature on the misinformation effect, how much people trust a source of information affects how much that source can mislead them, then we should expect to see a positive relationship between subjects ratings of trust and the difference between control and misled items. Put simply, the more people trust the robot witness, the larger the misinformation effect should be. But, that is not what we found. Instead, we found that trust was only trivially related to the size of the misinformation effect, and in the wrong direction; the more people trusted the robot, the smaller the difference between the control

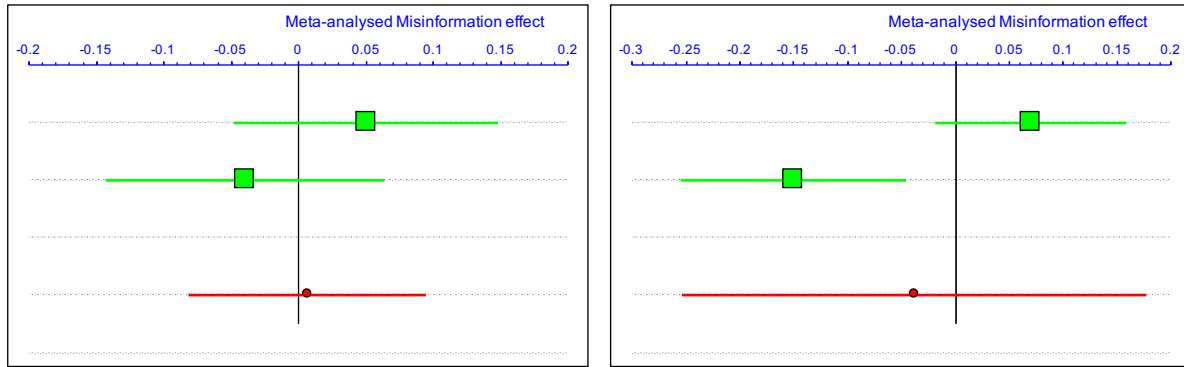


Figure 7. Forest plots for the mini meta-analyses comparing Android to Morph (left panel) and Robot to Morph (right panel). The x-axis is the size of the misinformation effect. The green lines refer to the difference between the means for Experiment 1 (top line) and Experiment 2 (bottom line). The red line refers to the meta-analysed difference. Error bars refer to 95% CIs.

and misled items on the memory test, $r = -0.02$, 95% CI $[-0.14, 0.10]$. Even when we separated the correlation by the source of the misinformation, we found that trust was only trivially related to the size of the misinformation effect, Android: $r = -.14$, 95% CI $[-0.35, 0.07]$, Morph: $r = 0.03$, 95% CI $[-0.18, 0.24]$, Robot: $r = -0.03$, 95% CI $[-0.24, 0.19]$.

Meta-analysis

In order to obtain a more precise estimate of the effect size, we combined the data from both experiments into two mini meta-analyses (Cumming, 2013). Using Exploratory Software for Confidence Intervals (ESCI: Cumming, 2013), we calculated two mini meta-analyses with the comparison between Android and Morph in the first meta-analysis, and Robot and Morph in the second meta-analysis. We display the forest plots for these results in Figure 7.

To interpret Figure 7, the x-axis can be conceptualised as the size of the misinformation effect, where positive numbers reveal a tendency to be more misled by the Morph, and negative numbers revealing a tendency to be more misled by the Android (left panel), or Robot (right panel). As the figure shows, when we consider the meta-analysed difference (as shown by the red line), it does not matter whether we consider the difference between the Android and Morph, $M_{\text{diff}} = -0.01$, 95% CI $[-0.08, 0.09]$, or the Robot and Morph, $M_{\text{diff}} =$

-0.04, 95% CI [-0.25, 0.18], the size of the effect is very small. In fact, the confidence intervals surrounding the meta-analysed difference in both cases contain positive and negative numbers, and zero, suggesting that there is plausibly no difference at all in how much each of the three robots we used were able to mislead people.

General Discussion

Our primary research question was: to what extent does a robot's appearance affect how likely it is to mislead people? Across two experiments we found that the answer to that question was not very much at all. In Experiment 1, the robot's appearance was only marginally able to influence how likely it was to mislead people. In Experiment 2, even when we drew attention to how much people trusted the robot eyewitness and allowed people to compare the different robots, people were still only marginally affected by how each robot looked. In fact, how much people trusted each robot was not plausibly related to how misled they were. When considered as a whole, the mini meta-analysis revealed that the robot's appearance had little to no effect on the size of the misinformation effect. These results suggest that even though our robot were able to mislead people, appearance did not matter.

But why did these robots' appearances not affect the size of the misinformation effect? One explanation is that there was not a large enough difference in how much people trusted each robot—people trusted the Morph only 5-6% less than the Android or Robot—and this difference might not have been enough to reveal any tendency to be misled by one robot more than another. Because the effect of misleading postevent information is large and robust, any manipulation designed to shift this effect needs to also be sufficiently robust to have any meaningful impact (Ullrich, Lewandowsky, & Tang, 2010). Future research should therefore address this issue by using materials that elicit a greater difference in trust ratings between the three robots. Although, it should be noted that, as shown in our second experiment, how much people trusted each robot was only trivially related to the size of the

misinformation effect. It is therefore possible that even though people trust certain robots less than others, they are still considered robots which possess the technology to be credible eyewitnesses. Trust might simply not be an important factor when assessing a robotic witness.

An alternative explanation, however, is that people quickly habituated to the appearance of the robot providing them with a witness statement and were therefore less likely to be affected by that robot's appearance. In fact, there is evidence to suggest that people do habituate to a robot's appearance (Fussell, Kiesler, Setlock, & Yew, 2008; Zlotowski et al., 2015). In one study, repeated interactions with a robot that fell into the uncanny valley was sufficient enough to reduce, or even eliminate, the feelings of unease that were initially associated with that robot (Zlotowski et al.). This finding suggests that, over time, people get used to a robot's appearance and can become less adversely affected by that appearance. In our study, we embedded an image of the robot eyewitness alongside each sentence of the witness statement. If people quickly habituate to the appearance of these robots, then by the time we start to introduce the misinformation, any difference in trust between these robots might already be eliminated. This idea would explain why we found little difference in how much each robot was able to mislead people. To address this issue, future research should only show the image of the robot at the start of the experiment. Doing so would reduce the potential for people to habituate to the robot. If we then see a difference in the size of the misinformation effect across robots, it might suggest that people in our study had habituated to the robot witness.

Our research is limited by the exclusion of one comparison—comparing the size of the misinformation effect between a human and robot eyewitness. In our study the Android is always called a robot, and so we do not know how much more or less a humanoid robot is able to mislead people compared to a human. A simple way to address this issue is to run an

experiment where one group of subjects are told that the android is an android, and another group of subjects are told the android is human. Comparing the size of the misinformation effect between these two groups could reveal the extent to which people can be misled by a robot, as well as how much this difference is affected by that robot's appearance.

Because the robots in our study were able to mislead people, our research has implications for research into how human memory is influenced by social sources. In our study, the robots were all equally able to mislead subjects regardless of their appearance. This finding stands in contrast to other published literature showing that the more credible or trustworthy a source appears, the more likely it is to mislead people (Dodd & Bradshaw, 1980; Zajac et al., 2013; French et al., 2011). In fact, it might be that the authority of a robot as an eyewitness outweighs any effects that the robot's appearance might have. It is therefore possible that when assessing the credibility of a source of information, people might weigh up not only factors such as how much they trust an eyewitness, but also how capable an eyewitness that source is. An eyewitness with high perceived capability might still possess the ability to mislead people, regardless of how much they are trusted.

But what does this research mean for robots in the field? Firstly, it means that the appearance of a robot does affect how we think and feel about it; a finding which replicates other studies demonstrating how people are affected by a robot's appearance (MacDorman & Ishiguro, 2006; Burleigh et al., 2013; Fussell et al., 2008; Prakash & Rogers, 2013).

Secondly, our research suggests that even though people feel more negatively about a robot that looks a particular way, people are still likely to believe the information that robot provides. Robots are becoming increasingly prevalent in our daily lives—from robots on a production line, to robots being introduced to our homes to serve as personal assistants—and trusting them is important given our increasing dependence on them. Robots have even been suggested to fill teaching positions in schools, particularly when it is difficult to find

substitute teachers. Our research suggests that even though people might initially find these uncanny valley robots eerie or creepy, this negative association should not affect how much people will believe the information these robots provide.

Human-machine interaction is not limited to just interacting with robots in a face-to-face setting. Recent developments have led to the widespread adoption of digital personal assistants on our smart phones such as: Siri (iOS), Cortana (Windows 10), or Google Now (Android). These personal assistants tend to speak in a mix of a computer synthesis of a human voice and an actual voice actor. The extent to which the voice used by these personal assistants affects how much we believe or trust the information they provide us is therefore an important issue. When information conveyed vocally by these personal assistants includes times of important meetings and appointments, or even how much medicine to take, mistrusting these personal/robot assistants can have real consequences.

In this study, we were primarily interested in the extent to which a robot's appearance can influence how much it can mislead people. Across two experiments we found that a robot's appearance had very little influence on its ability to mislead people, even when we drew attention to how much people trusted the robot eyewitness. Our research provides an important initial step in an under researched field—how our memory is influenced by a robot's appearance. Blindly trusting a robot can have adverse consequences, especially when high stakes information is involved, such as details of a crime. Due to both the malleability of human memory, and the fallibility of software/hardware, we should therefore be skeptical of the information robots provide us, and not simply believe our new robot overlords.

References

- Assefi, S. L., & Garry, M. (2003). Absolut® Memory Distortions Alcohol Placebos Influence the Misinformation Effect. *Psychological Science*, 14(1), 77-80. <http://dx.doi.org/10.1111/1467-9280.01422>
- Bethel, C. L., Eakin, D., Anreddy, S., Stuart, J. K., & Carruth, D. (2013, March). Eyewitnesses are misled by human but not robot interviewers. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on* (pp. 25-32). IEEE. <http://dx.doi.org/10.1109/hri.2013.6483498>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, 6(1), 3-5. <http://dx.doi.org/10.1037/e527772014-223>
- Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, 29(3), 759-771. <http://dx.doi.org/10.1016/j.chb.2012.11.021>
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge. <http://dx.doi.org/10.4324/9780203807002>
- Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, 19(6), 695-704. [http://dx.doi.org/10.1016/s0022-5371\(80\)90379-5](http://dx.doi.org/10.1016/s0022-5371(80)90379-5)
- 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. <http://dx.doi.org/10.1016/j.actpsy.2010.10.009>



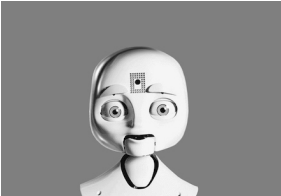





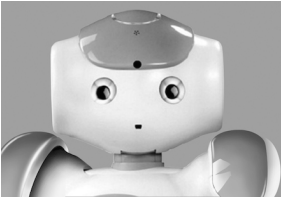

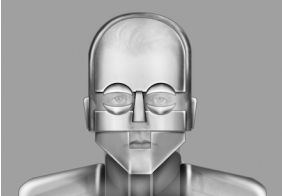

- Frumkin, L. (2007). Influences of accent and ethnic background on perceptions of eyewitness testimony. *Psychology, Crime & Law*, 13(3), 317-331. <http://dx.doi.org/10.1080/10683160600822246>
- Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, 42(1), 120-133. <http://dx.doi.org/10.1002/ejsp.862>
- Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008, March). How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 145-152). ACM. <http://dx.doi.org/10.1145/1349822.1349842>
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review*, 19(5), 847-857. <http://dx.doi.org/10.3758/s13423-012-0296-9>
- Greifeneder, R., Bless, H., & Pham, M. T. (2010). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*. <http://dx.doi.org/10.1177/1088868310367640>
- Kühnen, U. (2010). Manipulation checks as manipulation: Another look at the ease-of-retrieval heuristic. *Personality and Social Psychology Bulletin*, 36(1), 47-58. <http://dx.doi.org/10.1177/0146167209346746>
- Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38(8), 1087-1100. <http://dx.doi.org/10.3758/mc.38.8.1087>

- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of experimental psychology: Human learning and memory*, 4(1), 19. <http://dx.doi.org/10.1037//0278-7393.4.1.19>
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361-366. <http://dx.doi.org/10.1101/lm.94705>
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297-337. <http://dx.doi.org/10.1075/is.7.3.03mac>
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1), 1-23. <http://dx.doi.org/10.3758/s13428-011-0124-6>
- Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy*, 7 (4) 33-35.(Translated by Karl F. MacDorman and Takashi Minato in 2005 within Appendix B for the paper Androids as an Experimental Apparatus: Why is there an uncanny and can we exploit it?. In *Proceedings of the CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (pp. 106-118).
- Mori, K. (2007). A revised method for projecting two different movies to two groups of viewers without their noticing the duality. *Behavior Research Methods*, 39(3), 574-578. <http://dx.doi.org/10.3758/bf03193028>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, 19(2), 98-100. <http://dx.doi.org/10.1109/mra.2012.2192811>
- Prakash, A., & Rogers, W. A. (2013, September). Younger and older adults' attitudes toward robot faces effects of task and humanoid appearance. In *Proceedings of the human*

- factors and ergonomics society annual meeting* (Vol. 57, No. 1, pp. 114-118). SAGE Publications. <http://dx.doi.org/10.1177/1541931213571027>
- Takarangi, M. K., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology*, 20(5), 583-590. <http://dx.doi.org/10.1002/acp.1209>
- Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, 14(4), 329-338. <http://dx.doi.org/10.3758/bf03202511>
- Vornik, L., Sharman, S., & Garry, M. (2003). The power of the spoken word: Sociolinguistic cues influence the misinformation effect. *Memory*, 11(1), 101-109. <http://dx.doi.org/10.1080/741938170>
- Zajac, R., Dickson, J., Munn, R., & O'Neill, S. (2013). Trussht me, I know what I sshaw: The acceptance of misinformation from an apparently unreliable co-witness. *Legal & Criminological Psychology*. <http://dx.doi.org/10.1111/lcrp.12032>
- Złotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Frontiers in psychology*, 6. <http://dx.doi.org/10.3389/fpsyg.2015.00883>

Appendix A

Table A1
Images of the four families of robots used in the norming study

Family	Robot Type		
	Android	Morph	Robot
Family 1			
Family 2			
Family 3			
Family 4			

Note: Images were obtained from Prakash and Rogers (2013)

Appendix B

An example of one of the versions of the narrative used in the experiments reported here. Subjects read this narrative one sentence at a time, with each sentence appearing alongside a photo of the robot to which the statement was attributed to (see Table 2).

One afternoon, Eric Lewis, an average looking (although dishonest) guy with brown hair, had a job at a house in the city. Eric pulled up to the house in his blue “AJ’s Electricians” van and parked in the driveway. After getting his tool belt and drill kit out of the van, Eric went to the front door and retrieved the key that was left for him from under a flowerpot.

He unlocked the door and walked into the house, where he found a note from the homeowner on the hallstand, next to a vase of flowers. After reading the note, Eric put it in his pocket and put his drill kit down. Although being an electrician could get boring, one thing Eric liked about his job was that he could be nosy. He walked into a bedroom on the right, past the bed with its bright pink bedspread, and over to the dresser. He noticed the dresser was covered with a navy cloth. Curious, Eric bent down to lift it up and investigate the drawers underneath. Seeing nothing very interesting, Eric checked out the items on the dresser, including a wooden jewellery box. He opened the box and removed a pair of earrings, which he inspected carefully and—thinking of his girlfriend’s upcoming birthday—slipped them into his pocket.

Eric headed down the hallway to the living area, picking up his drill on the way. He thought the house was getting a little hot and stuffy, so he opened the French doors in the lounge and went into the adjacent kitchen. He put his drill kit down on the bench and thought he was feeling hungry, so he helped himself to a can of pepsi from the fridge and an apple from the fruit bowl. He continued to look around the kitchen, finding nothing interesting in

the pantry. Always curious about the secret lives of his customers, Eric rummaged through a pile of papers next to a mug on the kitchen bench, but there wasn't much interesting there either.

Thinking he should get down to work, Eric walked over to the broken oven and bent down to examine it. He removed a screwdriver from his tool belt and repaired the front panel. His kitchen repair finished, Eric headed down the hallway again to the second bedroom. In the bedroom, he kneeled down to check one of the power points that the homeowners said was not working. Once that was done, Eric got up to have a look around the room. He tried on a blue cap and checked his reflection in the mirror, but he didn't really like the way it looked on him. After replacing the cap on the bed, he browsed through the wardrobe, but didn't see anything he liked. Eric sat down on the bed to read a news magazine, but found it boring, and tossed it back on the floor. Looking up, he spotted a silver ring next to the stereo and tried it on. Thinking it looked pretty good on him, he pocketed the ring and started to look through a pile of CDs on the stereo. After selecting a CD he knew he would enjoy, Eric went back into the lounge to play it on the stereo in there, thinking he would listen to it while he finished his work in the lounge. His last job was a light fitting on the lounge wall. After removing the light cover, and working on the wiring, Eric flicked the switch on and off but nothing happened. He adjusted things a bit and the light finally came on.

Now that his work was completed, Eric decided that he deserved to relax for a while. He sat down on the couch, turned off the stereo and—finding the black remote on the coffee table in front of him—switched on the TV. He picked up a red photo album that was lying on the wooden coffee table and flicked through it. After a while, he checked his watch. Realizing he needed to get to his next job, Eric got up and turned the TV off. He retrieved the CD from the stereo and, thinking it would be good to listen to back in the van, he put it in his drill kit. After shutting and locking the French doors, he stopped to look at a picture on the wall. On

his way out, Eric decided to have a quick look through the bathroom cabinet, where he pocketed some prescription pills that he thought he might be able to sell. By now, he was late for his next job, so Eric hurried to the door and closed it behind him as he left.

Appendix C

The counterbalancing used in Experiments 1 and 2.

Table A2

Counterbalancing for Experiment 1

Counterbalance	Movie Version	Narrative	Robot	Family
1	1	1	Android	2
2	1	2	Android	2
3	1	1	Morph	2
4	1	2	Morph	2
5	1	1	Robot	2
6	1	2	Robot	2
7	2	1	Android	2
8	2	2	Android	2
9	2	1	Morph	2
10	2	2	Morph	2
11	2	1	Robot	2
12	2	2	Robot	2
13	1	3	Android	3
14	1	4	Android	3
15	1	3	Morph	3
16	1	4	Morph	3
17	1	3	Robot	3
18	1	4	Robot	3
19	2	3	Android	3
20	2	4	Android	3
21	2	3	Morph	3
22	2	4	Morph	3
23	2	3	Robot	3
24	2	4	Robot	3

Note: For Experiment 2, we removed one family of robots so the counterbalances were reduced to 12

Appendix D

The memory test used in Experiments 1 and 2. Each question appeared on a separate page with the corresponding confidence question appearing below the test question.

You will now be asked some questions about the video you saw. We are testing your memory for this video.

Each question has two parts:

- 1) the first part asks you about a particular item from the video;
- 2) the second part asks you how confident you are about your answer.

Here is a sample question.

Eric was working in _____

- a. a house b. a shop

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

WHEN YOU HAVE READ AND UNDERSTOOD HOW TO ANSWER THESE QUESTIONS, CLICK NEXT TO BEGIN THE TEST.

1. Eric was wearing _____

- a. overalls b. jeans

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

2. Eric ate _____

- a. an apple b. a banana

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

3. The magazine that Eric read was _____

- a. Time b. Newsweek

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

4. Eric read the note from the homeowner in the _____

- a. kitchen b. hallway

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

5. The tool that Eric used in the kitchen was _____

- a. pliers b. a screwdriver

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

6. In the lounge the picture Eric looked at was the _____ Tower

a. Eiffel b. Leaning

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

7. The bed in the first bedroom was _____

a. made b. unmade

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

8. In the second bedroom, Eric tested a _____

a. power point b. light fitting

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

9. Eric played a _____

a. video b. CD

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

10. In the second bedroom, Eric tried on a _____ cap

- a. blue b. black

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

11. The name of Eric's company was _____

- a. AJ's Electricians b. RJ's Electricians

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

12. Eric checked the time _____

- a. on his watch b. on the wall clock

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

13. The jewellery that Eric stole in the first bedroom was _____

- a. earrings b. a necklace

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

14. In the lounge Eric looked through a _____

- a. journal b. photo album

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

15. Eric's van was _____

a. blue b. red

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

16. Eric found the house key under a _____

a. door mat b. flower pot

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

17. Eric rummaged through papers that were next to a _____ mug

a. yellow b. white

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

18. Eric drank a can of _____

a. coke b. pepsi

How confident are you that your answer is correct?

1	2	3	4	5
Not at all				Very
Confident				Confident

19. In the bathroom Eric stole _____

a. pills

b. perfume

How confident are you that your answer is correct?

1
Not at all
Confident

2

3

4

5
Very
Confident

20. Eric stole _____ in the second bedroom

a. money

b. a ring

How confident are you that your answer is correct?

1
Not at all
Confident

2

3

4

5
Very
Confident
