The virtual maze: A behavioural tool for measuring trust

Joanna Hale, Madeleine E. M. Payne, Kathryn M. Taylor, Davide Paoletti and Antonia F. De C. Hamilton

University College London

Author Note

Correspondence should be addressed to Joanna Hale, UCL Institute of Cognitive Neuroscience, Alexandra House, 17 Queen Square, London, WC1N 3AR, UK. Email jo.hale.13@ucl.ac.uk. Telephone +44 (0) 20 7679 1126.

Madeleine E. M. Payne, email maddyempayne@gmail.com.

Kathryn M. Taylor, email katie.taylor16@hotmail.co.uk

Davide Paoletti, email d.paoletti@ucl.ac.uk

Antonia F. de C. Hamilton, email a.hamilton@ucl.ac.uk

Acknowledgement

This work was supported by the MoD through Dstl under Grant DSTL/PS79473 and the ERC under Grant 313398-INTERACT.

Abstract

Trusting another person may depend on our level of generalised trust in others, as well as perceptions of that specific person's trustworthiness. However, many studies measuring trust outcomes have not discussed generalised versus specific trust. To measure specific trust in others, we developed a novel behavioural task. Participants navigate a virtual maze and make a series of decisions about how to proceed. Before each decision, they may ask for advice from two virtual characters they have briefly interviewed earlier. We manipulated the virtual characters' trustworthiness during the interview phase and measured how often participants approached and followed advice from each character. We also measured trust through ratings and an investment game. Across three studies we found participants followed advice from a trustworthy character significantly more than an untrustworthy character, demonstrating the validity of the maze task. Behaviour in the virtual maze reflected specific trust rather than generalised trust, whereas the investment game picked up on generalised trust as well as specific trust. Our data suggests the virtual maze task may provide an alternative behavioural approach to measuring specific trust in future research, and we demonstrate how the task may be used in traditional laboratories.

Keywords: trust, nonverbal behaviour, investment game, virtual reality, behavioural measure

In everyday life, we often have to weigh up how much we trust strangers or people we have only briefly met. Can I trust that passer-by to direct me to the station? If I lend the new intern my stapler, will I get it back? In some situations, such as criminal investigations, there are high stakes attached to the decision whether or not to trust someone. In this paper we consider how to measure the level of trust one person feels towards a specific stranger and develop a new behavioural task which may be used to test social factors that influence levels of trust.

The measurement of trust is a vast and complicated topic, spanning domains in psychology, neuroscience, sociology, behavioural economics and organisation science (Ashraf, Bohnet, & Piankov, 2006; Bachmann & Zaheer, 2006; Blomqvist, 1997). Across these fields there are many different definitions of trust (Ashraf et al., 2006; Bachmann & Zaheer, 2006; Ben-Ner & Halldorsson, 2010), but very broadly speaking there are two main ways we can think about an individual's level of trust towards someone else. On one hand, we can treat their level of trust as a stable personal characteristic, a reflection of how much they trust others in general. This is often termed 'generalised trust' (Couch & Jones, 1997; Freitag & Traunmüller, 2009). On the other hand, we can treat their level of trust as a specific reaction to the other person, perhaps based on having a close relationship with them, or other social cues if the person is a stranger. The term 'interpersonal trust' is often used to refer to this kind of trust between people in a close relationship (Johnson-George & Swap, 1982; Rotter, 1967, 1971); we will use the term 'specific trust' to cover strangers as well.

In this paper, we are interested in how to measure specific trust towards one particular person. We will begin by reviewing three major methods for trust measurement available to social psychologists and social neuroscientists, considering the suitability of each method. The first and oldest method is to administer questionnaires that ask people to self-report how much they trust others (Johnson-George & Swap, 1982; Rempel, Holmes, & Zanna, 1985;

Rosenberg, 1957; Rotter, 1967). A more recent behavioural approach is to measure how much money an individual will entrust to another player in an economic game (Berg, Dickhaut, & McCabe, 1995; Glaeser, Laibson, Scheinkman, & Soutter, 2000; Tzieropoulos, 2013). Finally, an alternative behavioural approach is to measure trust in terms of willingness to ask for and endorse information from an informant (Clément, Koenig, & Harris, 2004; Harris & Corriveau, 2011). The first two approaches are the most widely used across psychology and other social sciences, but we argue the 'ask and endorse' approach is most suitable for measuring trust towards specific strangers. We then introduce our novel method for measuring trust using a virtual maze task, which builds on some of the advantages of ask-endorse paradigms.

Self-report questionnaires

For many decades, researchers have used self-report questionnaires to gauge levels of trust. Most of these have been designed to measure generalised trust (Couch & Jones, 1997; Freitag & Traunmüller, 2009; Glaeser, Laibson, Scheinkman, & Soutter, 1999) or interpersonal trust towards a close interaction partner such as a spouse or family member (Couch & Jones, 1997; Johnson-George & Swap, 1982; Rotter, 1967, 1971; Sorrentino, Holmes, Hanna, & Sharp, 1995). In comparison, there are very few validated scales that capture specific trust towards a stranger during an experiment. One exception is McCroskey & Teven's (1999) trustworthiness scale in which participants rate six dimensions of trustworthiness, in line with research highlighting the multidimensional nature of trustworthiness judgments and other social impressions (e.g. Ben-Ner & Halldorsson, 2010; Fiske, Cuddy, & Glick, 2007). Other researchers have often used just one or two items to measure the perceived trustworthiness of a target person (e.g. Maddux, Mullen, & Galinsky, 2008; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). For example, participants might rate the target from 'not at all' to 'extremely' trustworthy (Willis &

Todorov, 2006), or make a yes/no judgement about the target's trustworthiness and back this up with a confidence rating (Todorov et al., 2009). Approaches like these have the advantage that they are straightforward to administer and therefore may translate well to settings outside of the laboratory, particularly to clinical contexts in which behavioural measures may impractical and time-consuming. However, one or two items may not fully capture the nature of trustworthiness (Ben-Ner & Halldorsson, 2010) and perceiving someone as trustworthy may not always equate to trusting them. It is also difficult to interpret and compare questionnaire results across studies, since the items used vary and are often specific to the experimental setting, e.g. 'how much did you trust the other party during the negotiation?' (Maddux et al., 2008). This can be an advantage for investigating context-specific trust towards different sources (e.g. political leaders vs. scientists on the topic of stem cell research; Liu & Priest, 2009) but is less suitable for social cognitive research into what makes people trust or not trust a specific stranger.

The lack of validated scales for measuring trust towards strangers may be due to concerns over the validity of self-report methods. Questionnaire items are open to interpretation, which may undermine the validity of responses. Trust questionnaires can be particularly susceptible to ambiguity due to the multiple meanings and interpretations of trust (Ben-Ner & Halldorsson, 2010; Lyon, Möllering, & Saunders, 2012), particularly across cultures (Hooghe, Reeskens, Stolle, & Trappers, 2009; Miller & Mitamura, 2003). Assuming the interpretation of items were unambiguous, participants still may not have accurate access to their internal feelings (Chan, 2009); even if they do, their self-reports may be biased by social norms or demand characteristics (McCambridge, de Bruin, & Witton, 2012). Finally, even if people report accurate internal feelings, these may be poor predictors of external behaviour (Armitage & Christian, 2003); for example, survey measures of generalised trust actually predict *trustworthy* behaviour better than trusting behaviour (Glaeser et al., 1999).

For these reasons, implicit behavioural measures are often preferred over explicit self-report ratings.

The investment game

A major behavioural method for investigating trust emerged from behavioural economics. This method is aligned with the view that trust involves a ‘voluntary transfer of a good or favour to someone else, with future reciprocation expected but not guaranteed’ (Gunnthorsdottir, McCabe, & Smith, 2002, p. 50). The method was developed by Berg, Dickhaut and McCabe (1995), who designed a simple investment game played between two people: an investor and a trustee. The investor was given $10 (different amounts have been used in subsequent studies; Johnson & Mislin, 2011) and had to decide how much of that $10 to send to the trustee, knowing that the amount they sent would be tripled before it was given to the trustee. Then the trustee had to decide how much of the tripled amount to return to the investor. The game measures trust behaviour in terms of the percentage of money the investor is willing to send to the trustee. Earlier trust games, such as the prisoners’ dilemma, typically required an all-or-nothing decision to trust or distrust the other player, which did not provide such a sensitive measure (Schniter, Sheremeta, & Shields, 2013). Thus, the investment game has come to dominate the field (Johnson & Mislin, 2011).

It is unclear to what extent the investment game taps into generalised trust versus specific trust. Although the investment game was originally designed to answer questions about generalised trust (Berg et al., 1995; Glaeser et al., 2000; McEvily, Radzevick, & Weber, 2012), several studies have found that people’s investments correlate poorly with generalised trust questionnaires (Ashraf et al., 2006; Glaeser et al., 1999) and relate instead to perceptions of the other player’s trustworthiness (McEvily et al., 2012). In psychology and neuroscience, the investment game has been used to test factors that may affect specific trust towards known and unknown trustees. For example, studies have found that participants make significantly higher investments when the trustee is happy (Tortosa, Lupiáñez, & Ruz, 2013; Tortosa, Strizhko, Capizzi, & Ruz, 2013), belongs to a racial in-group (Stanley et al., 2012) or coordinates their nonverbal behaviour with the investor (Launay, Dean, & Bailes, 2013; Verberne, Ham, & Midden, 2015; Verberne, Ham, Ponnada, & Midden, 2013). On the other hand, investment behaviour is also found to correlate with traits such as altruism (Ashraf et al., 2006; Ben-Ner & Halldorsson, 2010; Cox, 2004) and risk-seeking (Karlan, 2005; McEvily et al., 2012; Schechter, 2007), and to vary according to cultural norms (Johnson & Mislin, 2011; Willinger, Keser, Lohmann, & Usunier, 2003). This suggests that even if the investment game is not correlated with self-reported levels of generalised trust, it is sensitive to stable individual characteristics which may be proxies of generalised trust. Overall, the amount someone invests is likely to reflect a mixture of generalised trust and specific trust towards the other player, but it is unclear how levels of these are weighted in different people and different versions of the investment game.

In our own experiments, we have had little success using the investment game to test specific trust towards interactive virtual characters (Hale, 2016). Across three studies with a total of 78 participants we have consistently found that how much people invest with a particular virtual character does not correlate with their rating of that character’s trustworthiness. Instead, we find that their investments towards two different virtual characters are highly correlated, suggesting that the investment game was picking up on a stable level of generalised trust. This motivated us to develop an alternative behavioural approach that would be more sensitive to levels of specific trust and less affected by generalised trust or other individual traits.

The investment game also has some practical limitations. Firstly, it is hard to know how far people’s investment behaviour in this abstract game may be used to infer how they would trust someone in the real world. In the investment game, participants have to make an explicit decision about the amount of money to send the trustee, but in real life people's trust decisions may be more implicit. Secondly, the rules of the investment game are somewhat complicated to explain, causing differences in how participants perceive and interpret the game (Macko, Malawski, & Tyszka, 2014). The task complexity means that the investment game may not be suitable for young children, or participant groups whose understanding is otherwise impaired. Furthermore, healthy adult participants may 'overthink' their response; for example, participants make more cautious investments when they have less time to make a decision (Tzieropoulos, Grave De Peralta, Bossaerts, & Gonzalez Andino, 2011) or know they will be paid randomly at the end of the study (Johnson & Mislin, 2011). Finally, the investment game is dyadic and over successive rounds players learn about each other's trust or trustworthiness (King-Casas et al., 2005). This means there is a one-shot opportunity on the first round to measure the investor's initial trust towards the trustee. It is unclear whether an averaged measure may be derived across multiple trials of the 'first round' where the investor does not find out the trustee's decision.

The ask-endorse paradigm

An alternative behavioural approach to measuring trust comes from developmental psychology. In order to investigate the extent to which children will trust information from a teacher or informant, Koenig et al. (Koenig, Clément, & Harris, 2004; Koenig & Echols, 2003; Koenig & Harris, 2005, 2007; Pasquini, Corriveau, Koenig, & Harris, 2007) developed a paradigm in which children implicitly had to consider the trustworthiness of two puppets or video characters. Having seen one puppet or video character give accurate information, and the other give inaccurate information, each child had to make two choices. Firstly, they had to choose which one they would ask in order to learn something new. Each puppet or character would give conflicting testimony. Then the child would have to choose which testimony they believed. This procedure was repeated over multiple trials. Thus the paradigm provides two measures of trust: firstly, whether the child would *ask* for information, and secondly whether they would *endorse* that information. These may be called ‘selective’ trust measures since trust is inferred from which informant the child selects. Other selective trust research by Mills et al. (Johnston, Mills, & Landrum, 2015; Landrum, Mills, & Johnston, 2013; Mills, Legare, Bills, & Mejias, 2010) has also measured how often children would ask questions to different puppet informants to help them figure out a puzzle. Studies using selective trust paradigms have shown that children tend to trust informants who are nice, smart and honest (Landrum, Eaves Jr, & Shafto, 2015; Lane, Wellman, & Gelman, 2013), as well as people who are attractive (Bascandziev & Harris, 2014) or belong to their in-group (Kinzler, Corriveau, & Harris, 2011; VanderBorght & Jaswal, 2009). Several of the effects found with children have been replicated in adults using the same approach (Landrum et al., 2015; Lane et al., 2013).

There are several advantages to this approach for measuring trust towards a specific stranger. Firstly, it represents an ecologically valid scenario, i.e. ‘based on limited experience of this person, will you ask them for advice and trust what they say?’ We might consider the same questions when asking a passer-by for directions, or conducting an investigative interview. Therefore, the ask-endorse scenario might be more representative of everyday trust decisions than the investment game scenario. Secondly, the decision to ask or endorse taps into implicit trust behaviour, which may provide a ‘purer’ estimate of trust levels than explicit measures. Thirdly, by framing each decision in terms of selecting one informant versus the other, participants have to use their perceptions of each informant’s trustworthiness and therefore the task should be more sensitive to specific trust than generalised trust.

The virtual maze

We aimed to design a novel behavioural task for measuring trust towards specific strangers, and decided to implement the task in virtual reality. As virtual reality technology becomes increasingly sophisticated and more widely available, it is becoming an increasingly popular tool in social psychology and social neuroscience (Blascovich et al., 2002; McCall & Blascovich, 2009). This is because virtual characters and virtual spaces can be manipulated in systematic ways that would be hard to achieve with confederates in a physical laboratory (Fox, Arena, & Bailenson, 2009). For example, it is easy to program a virtual character that subtly blinks or nods at a certain rate (Bailenson & Yee, 2005; Gratch et al., 2006), or a virtual space which is much larger than the average laboratory. Many studies have shown that participants react to virtual scenarios as they would in real life (Durlach & Slater, 2000; Fox et al., 2009; Garau, Slater, Pertaub, & Razzaque, 2005), so virtual reality also offers an opportunity to closely replicate everyday situations under controlled conditions. As well as affording high ecological validity, there are also novel opportunities to measure implicit behavioural responses such as where participants direct their gaze in the 3D space, or how closely they will approach a virtual character (Khooshabeh et al., 2011; McCall, Blascovich, Young, & Persky, 2009).

Exploiting some of the strengths of virtual reality, we designed a task where participants navigate through a virtual maze and may choose to trust virtual characters about which way to proceed. Participants find themselves in a virtual maze made up of a series of identical rooms connected by corridors. Each time they enter a new room, they face two doors and must make a decision about which door to proceed through. To help them decide, they may approach two virtual characters for advice, although they do not have to. When approached, each virtual character will indicate which door they think is the one to take. The participant keeps making decisions until they are told that they have found the way out of the maze. In fact, there are no right or wrong choices about which way to go. Instead, we are able to randomly generate endless rooms and corridors until the participant has gone through a specified number of rooms (trials) and we tell them they have found the way out. At the end, we can measure how often the participant approached each character for advice, and how often they followed advice they received from each character. The virtual maze therefore follows a similar approach to the ask-endorse paradigm, although the virtual maze task is more implicit in that participants are not prompted on every trial which character they want to ask for advice, but are instead left to make an implicit choice to approach neither, one or both characters.

In this paper, we present three studies in which we piloted the virtual maze task and explored how it compares to other trust measures. Firstly, we aimed to test whether people's decisions in the virtual maze are sensitive to differences in trustworthiness between two virtual characters. Therefore, in the first two studies participants got to know two different virtual characters through a short interview where the participant asked each character some prepared questions. During the interview, we manipulated the trustworthiness of each character through their verbal answers and nonverbal and vocal behaviour. The manipulation was designed to achieve a large effect size, since we did not know the sensitivity of the virtual maze task. We predicted that participants would decide to approach and follow the advice of a trustworthy character significantly more often than an untrustworthy character in the virtual maze. In order to compare our task with major alternative methods, in Studies 1 and 2 we also included ratings of each character's trustworthiness and an investment game. We further aimed to explore the extent to which each measure showed a correlation between the two virtual characters. If trust in one character is correlated with trust in the other character, that would indicate the level of trust towards each one is being driven by stable individual differences among participants, i.e. generalised trust. On the other hand, if a measure shows no correlation between two characters that suggests it is sensitive to specific trust rather than generalised trust. In the third study, we implemented a purer manipulation of trustworthiness by using the investment game as a manipulation instead of a dependent measure, following two studies by Franzen et al. (Franzen et al., 2011; Lis et al., 2013). In Study 3, we also programmed a low-tech version of the virtual maze task on a standard computer screen to demonstrate how the task could be used in standard laboratories without VR equipment or software.

Study 1

Methods

Participants. Twenty participants (13 female) were recruited through email advertisements to the Institute of Cognitive Neuroscience (ICN) departmental participant pool. A power calculation using G*Power showed that 20 participants would be sufficient to detect a large effect size (d_z = 1.1) with power of 0.98 or a medium effect size (d_z = 0.71) with power of 0.91. All participants gave written informed consent and received payment of £7, plus a bonus of up to £3 depending on how they played the investment game. The study was granted ethical approval by the MoD Research Ethics Committee and the ICN Ethics Chair.

Virtual reality system. Participants sat at a desk in front of a 90 x 160cm projector screen. We used Vizard virtual reality software (Worldviz, 2014) to display a virtual environment on the screen (Figure 1A). During the interview phase of the experiment, the virtual environment looked like an extension of the physical desk and walls of the laboratory. We programmed life-sized virtual characters to appear seated on the other side of the desk, facing the participant. We programmed a virtual barrier to occlude the virtual character at the end of the interview phase. Instructions and stimuli for the following tasks were displayed on the virtual barrier. The virtual maze task was also completed on the projector screen, but the maze environment was not designed to look like an extension of the laboratory. Instead, the participant saw virtual corridors and rooms similar to playing a video game.

Virtual characters. We prepared two male virtual characters for the experiment, named Mike and Ryan. The characters' appearances (Figure 1B) were selected from a collection of characters ('Complete Characters HD') supplied by Rocketbox Libraries for Vizard. We scripted everything the characters said during the experiment, which we pre-recorded from two male volunteers with native British accents. The pre-recorded speech was triggered by the experimenter or the computer program. The characters were programmed to move their jaw according to the amplitude of the pre-recorded speech, so that it looked like they were speaking. Audio speakers were hidden behind the projector screen so that the sound of the character's voice came from their virtual location.

Trustworthiness manipulation. We manipulated the trustworthiness of the two virtual characters during the initial interview phase of the experiment. Mike was designed to seem trustworthy, whereas Ryan was designed to seem untrustworthy. We achieved the manipulation through verbal, non-verbal and vocal signals.

Verbal signals. Mike and Ryan gave different scripted responses to three interview questions: (1) *What is your occupation?* (2) *What did you do last weekend?* (3) *What are your plans for the summer?* Mike made statements demonstrating reliability, e.g. *'I promised to raise £800 in sponsorship and I managed to smash the target'*, whereas Ryan indicated irresponsibility, e.g. *'things didn't go so well at my last job and I basically ended up getting fired after missing too many deadlines'*. For full scripts see Appendix.

Non-verbal signals. We programmed Mike and Ryan to make different amounts of eye contact, which people often use as a nonverbal signal to trustworthiness (even if it is not a reliable signal; DePaulo et al., 2003). Mike averted his gaze at random intervals of 6-9s, whereas Ryan averted gaze at random intervals of 3-6s and consequently made less eye contact. Both characters looked away at a random location for 0.75-1.75s before returning to make eye contact.

Facial appearance is also used as a nonverbal cue to trustworthiness (Todorov, 2008; Todorov & Duchaine, 2008). We obtained ratings of trustworthiness on a 1 to 6 scale for six different virtual characters by 52 participants in an online questionnaire. The appearance rated the most trustworthy (M = 3.89, SD = 1.15) was assigned to Mike and the appearance rated as least trustworthy (M = 2.79, SD = 1.16) was assigned to Ryan (c.f. Ewing, Caulfield, Read, & Rhodes, 2014). There was a significant difference in trustworthiness ratings between the two characters ($t(51) = 5.37$, $p < .001$).

All other nonverbal behaviour (e.g. posture, head tilting) was controlled using the same idle animation for both characters.

Vocal signals. Vocal hesitations and disfluencies are sometimes seen as a signal of untrustworthiness (Anderson, DePaulo, Ansfield, Tickle, & Green, 1999; Hosman & Wright, 1987). Therefore, the volunteer voicing Mike was instructed to speak clearly without many hesitations, whereas the volunteer voicing Ryan was instructed to mumble and make regular hesitations (e.g. *'umm'*).

Virtual Maze Task.

Virtual maze environment. The virtual maze was generated from a series of identical rooms (Figure 1C). The rooms were connected by twisting sections of corridor designed to enhance the illusion of being in a maze. The participant entered each room through a brown door; at the far end of the room, there was a red door and a blue door. In each room there were also two semi-transparent ‘hologram chambers’, where the virtual characters appeared as holograms from outside the maze. Mike always appeared on the left and Ryan always appeared on the right, but participants did not show any preference for approaching the left versus right hologram chamber (see Appendix). Whenever the participant got close to a hologram chamber, there was a sound effect and the chamber became more transparent. At the same time, the character inside the chamber would spin to face the participant and deliver some verbal advice. Participants were able to navigate the virtual environment by using a keypad to turn the viewpoint and move forwards or backwards.

Trial procedure. Going through one room corresponded to completing one trial. Each participant completed twelve trials in total. In each trial, they had to make a choice about whether to proceed through the blue door or the red door. The participant was able to approach neither, one or both characters to receive advice about which door to choose, although this was not explicitly instructed. If approached, the characters randomly delivered uncertain advice about which door the participant should choose (e.g. *‘It’s blue this time, I think’*). There was no ‘correct’ door on each trial. Instead, the maze was completed after twelve trials in which the participant approached at least one character. If a participant asked neither character, that trial was recorded but did not count towards the requisite twelve trials. This ensured we had twelve trials in which the participant received some advice about which way to go, thus providing data about how much they trusted that advice.

Character advice. Mike and Ryan were programmed to advise the red door in half the trials and the blue door in the other half. They were also programmed to advise the same door as each other in half the trials and different doors in the other half. In order to generate the verbal advice stimuli, we pre-recorded twelve scripted phrases. Then we paired the phrases in order to create twelve combinations of advice stimuli (Table 1). The order of the stimulus combinations was randomised for each participant. Note that the participant would only receive an advice stimulus if they approach a character for advice. Therefore, some participants may not have received both parts of every stimulus combination.

(Table 1 about here)

Dependent measures. We measured three dependent variables on each trial of the maze task. Firstly, we recorded whether each character was approached for advice. Secondly, we recorded which character the participant approached first (if the participant only approached one character, then that character was treated as being approached first). Thirdly, we recorded whether the participant followed the advice of each character. We averaged each of these variables over the total number of trials completed, giving us (1) the proportion of trials on which each character was approached for advice; (2) the proportion of trials on which each character was the first (or only) character approached for advice; and (3) the proportion of trials on which the participant followed each character's advice.

Procedure.

Interview. Participants were instructed that they were going to interview an individual virtual character called Mike (or Ryan) in order to get to know him. The participant was given a sheet with three prepared questions, and instructed to ask each question one at a time during the interview. At the start of the interview, Mike introduced himself and prompted the participant to begin asking questions. When the participant asked a question, the researcher triggered a pre-recorded scripted answer. The interview lasted around 5 minutes. At the end of the interview a virtual barrier appeared to occlude Mike and present the next tasks.

Character ratings. Participants rated their agreement with 10 statements about Mike by clicking on a continuous scale from 0 (strongly disagree) to 1 (strongly agree). Each statement began *‘I think Mike is...’* followed by five items measuring rapport (*likable/engaging/kind/unfriendly/unpleasant*) and five items measuring trustworthiness (*trustworthy/honest/responsible/unreliable/insincere*). The statements were presented in a randomised order. We reversed the scores for negatively valenced items and averaged the responses to provide one score for rapport (α =.86) and one score for trustworthiness (α = .76).

Investment Game. Participants completed five trials of the Investment Game with Mike, based on Berg et al.’s (1995) paradigm. At the start of each trial the participant had £1 to play with (this did not accumulate over trials). They could invest any proportion of the £1 with Mike. Mike would always triple their investment. Then he would choose a proportion of the tripled amount and return it to the participant, but his decision was not revealed. The participant would end the trial with any money they chose not to invest, plus any money that Mike chose to return them. Unknown to the participant, we programmed the task so that 50% of the tripled money was always returned (i.e. a maximum of £1.50 from Mike or Ryan). We told participants at the start of the game that one trial would be selected at random and they would find out the outcome of that trial, which would be paid as a cash bonus. To measure how much the participant trusted Mike, we measured the proportion of £1 that the participant chose to invest as an average across the five trials.

Once the participant had completed the interview, character ratings and investment game with Mike, then they completed the same three steps with Ryan. The order in which participants met Mike and Ryan was counterbalanced.

Virtual Maze. After interacting separately with Mike and Ryan, the participant completed the virtual maze task. First the participant practiced using a keypad to navigate around a virtual space and approach hologram chambers in order to receive a greeting message from Mike or Ryan. After the practice, the participant was instructed that their task was to find the way out of the virtual maze as quickly as possible. They were also told ‘There are some rooms in the maze where you will have to make a choice about which way to go. To help you decide, Mike and Ryan will be able to give you remote advice from outside the maze. They will appear as holograms in the room and you can go up to them to get advice’. Then participants began the task. After completing twelve trials, they were told they had successfully completed the maze.

Finally, participants completed a questionnaire unrelated to the current study, and then received payment. The experiment took approximately 40 minutes.

Results and Discussion

Missing data. One participant completed only five trials of the virtual maze due to motion sickness, but their data was included in the analyses. Three other participants briefly paused during the virtual maze task due to feeling motion sick, but this was not enough to cause concern that the task was too unpleasant.

Trust towards Mike and Ryan. We carried out paired-samples *t*-tests to determine whether there was a significant difference between Mike and Ryan on each of our six dependent measures: rapport rating, trustworthiness rating, percentage investment, approaching for maze advice, approaching first, and following maze advice. We applied a Bonferroni correction for multiple comparisons. Mike scored higher than Ryan on every measure, indicating participants liked and trusted Mike more than Ryan (Table 2). In the virtual maze participants approached Mike significantly more often than Ryan, although they did not approach Mike first significantly more often than they approached Ryan first. Furthermore, participants followed advice from Mike significantly more than Ryan. These results show that the maze task was sensitive to our manipulation, although it is possible that the effects were due to differences in the likeability of each character, rather than their trustworthiness.

(Table 2 about here)

Correlations between Mike and Ryan. For each measure, we examined the correlations between scores for Mike and Ryan (Table 3). Investment was the only measure which showed a significant correlation between the two characters, which was strongly positive. This means that participants were fairly consistent in their investments towards the two characters, suggesting investment behaviour might have been driven by generalised rather than their perception of the specific characters' trustworthiness. However, we cannot tell whether this evidence for generalised trust may be explained by variability in traits such as risk aversion and altruism (McEvily et al., 2012), or whether different participants simply interpreted the task instructions differently (Macko et al., 2014).

(Table 3 about here)

Summary. We found that the maze task was sensitive to our trustworthiness manipulation, showing a significant difference in how often people would approach and follow the advice of the trustworthy virtual character (Mike) versus the untrustworthy character (Ryan). Our manipulation led to large effect sizes on the measures we took from the virtual maze, as well as our other dependent measures. The investment game was the only measure for which people's responses towards Mike and Ryan were significantly correlated. This suggests that people's investments may reflect a generalised level of trust, or other stable individual characteristics.

Study 2

In Study 2, we aimed to replicate Study 1 using a more controlled manipulation of trustworthiness, where each character would be matched in likeability. We also aimed to adapt the maze task for use with a head-mounted display (HMD), and included an extra questionnaire to measure how real the virtual experience seemed to participants and record any feelings of motion sickness associated with the HMD.

Methods

Participants. Twenty-four participants (17 female, M_{age} = 25.9, SD_{age} = 10) were recruited through email advertisements to the ICN departmental participant pool. The sample size was based on Study 1, which found large effects with a sample of 20. All participants gave written informed consent and received payment of £7 per hour, plus a bonus of up to £3 depending on how they played the investment game. The study was granted ethical approval by the MoD Research Ethics Committee and the ICN Ethics Chair.

Virtual reality system. We used Vizard virtual reality software (Worldviz, 2014) to display virtual environments in an Oculus Rift DK2 head-mounted display (HMD). This device allows people to look around a virtual 3D space as if they are really there. Participants wore the HMD while seated at a physical desk in our lab. During the interview phase of the experiment, participants saw a virtual desk in place of the physical desk, inside a virtual room that looked like a typical psychology laboratory. We programmed a virtual screen to occlude the virtual character at the end of the interview phase and display the next tasks to the participant. The virtual maze task was also displayed via the HMD. Participants wore stereo headphones throughout the experiment to hear the characters speaking and other sound effects. They were provided with a joystick to make responses during the tasks and navigate through the virtual maze.

Trustworthiness manipulation. The Mike and Ryan characters from Study 1 were slightly modified for Study 2. In this study, Mike was designed to seem reliable and Ryan was designed to seem unreliable but we aimed to make both characters equally likeable. Therefore, we scripted new questions and responses for the characters to deliver in the interview phase. We altered verbal, non-verbal and vocal signals of trustworthiness from each character as follows:

Verbal. Mike and Ryan delivered new scripted responses to seven interview questions. Mike made statements demonstrating reliability, e.g. *'I ended up graduating with the highest grade'*, whereas Ryan indicated unreliability, e.g. *'I only had a few lectures a week and I used to miss them all the time'*. Both characters showed their likeability, e.g. *'I met so many people that I'm still really close with'* (Mike); '*I also get along with everyone'* (Ryan). For full scripts see Appendix. To validate the new scripts, a separate pilot sample provided online ratings of rapport and trust towards the speakers. We presented the scripts from Study 1 and Study 2 as if they were transcripts from real interviews (in the scripts from Study 1 we changed the characters' names to David and Ben). Pilot participants rated significantly more trust ($t(14) = 8.14$, $p < .001$, $d = 3.73$) and rapport ($t(14) = 8.46$, $p < .001$, $d = 2.16$) towards Mike than Ryan in the Study 1 scripts (Table 4). In the new scripts for Study 2 (Table 4), pilot participants rated significantly greater trust towards Mike ($t(14) = 3.56$, $p = .003$, $d = 1.48$) but there was no significant difference in rapport towards Mike and Ryan ($t(14) = .67$, $p = .514$, $d = 0.24$), supporting the validity of our manipulation. However, it should be noted these pilot ratings were based on a small sample size.

Table 4 about here.

Non-verbal. As in Study 1, Mike and Ryan were programmed to make different amounts of eye contact. In this study, we also manipulated the characters' promptness as a signal of reliability. At the start of the interview phase, the participant was asked to wait while the current virtual character got ready. Mike was ready after a short delay of 4 seconds, saying '*Ok, yep, I'm ready to start!',* whereas Ryan took 14 seconds to get ready before saying 'S*orry I'm late. Yeah, OK, I'm ready now'.* All other nonverbal behaviour was controlled using the same idle animation for both characters.

Vocal. As in Study 1, the volunteer voicing Mike was instructed to speak without many hesitations, whereas the volunteer voicing Ryan was instructed to make more hesitations. However, both volunteers spoke clearly and engagingly so as to avoid differences in friendliness or likeability.

Virtual maze task. We slightly adapted the virtual maze task from Study 1 for use with the HMD. Instead of using a keypad to navigate the virtual maze, participants used a joystick. We triggered extra sound effects when the participant went through doors in the virtual maze, in order to increase realism. We also rendered a plainer texture on the virtual walls and simplified the sections of corridor joining each room in order to reduce motion sickness associated with navigating narrow and twisty sections of corridor. All other aspects of the virtual maze task remained the same.

VR Questionnaire. We included a short questionnaire about the participant's experience in immersive virtual reality. Participants indicated their agreement with statements on scale from 1 (strongly disagree) to 7 (strongly agree). Four statements assessed how much the virtual world was real ('presence'), e.g. *'During the experience, the interview felt like the real world for me'*. Four statements assessed how much the virtual characters seemed real ('co-presence'), e.g. *'My feelings and emotions in relation to Mike/Ryan were as if they were real'*. Items were based on presence and co-presence items used in other virtual reality studies (Friedman et al., 2014; Pan, Gillies, Barker, Clark, & Slater, 2012). The questionnaire also asked participants to indicate whether they felt any motion sickness, queasiness, headache or eye strain whilst wearing the HMD.

Procedure. Each participant was seated at a desk and given verbal instructions from the experimenter. Then they were fitted with the HMD and saw themselves in a virtual laboratory. They could see a laptop on a desk in front of them, but a large screen displaying instructions blocked the participant’s view of the room on the other side of the desk (Figure 2A). The participant was given a few minutes to become accustomed to the virtual environment. They also practiced using the joystick to trigger instructions and log responses. When the participant had completed the practice, they were instructed they were going to interview a character called Mike and were asked to press a button on the joystick when they were ready to start the interview. Then the participant was instructed to wait while Mike got ready. After a specified delay, the participant heard a door opening and a chair moving somewhere on the other side of the large screen, and then Mike said he was ready. At that point, the screen moved up to the ceiling so that the participant could see Mike (Figure 2B). Interview questions were displayed to the participant one at a time on the laptop. When the participant asked each question, the researcher triggered a pre-recorded scripted answer from Mike. The interview lasted around five minutes. At the end of the interview, the large screen was lowered again to occlude Mike and present the next tasks.

Character ratings. Immediately following the interview, participants rated their feelings of rapport and trust towards Mike, as in Study 1.

Investment Game. Next, participants completed the investment game as in Study 1. Data from Study 1 indicated that participants made highly consistent choices over five trials of the investment game. Therefore in Study 2 we decided to reduce the number of trials to one, consistent with the traditional paradigm (Berg et al., 1995). Reducing the number of trials also reduced the time participants spent wearing the HMD.

Once participants had completed these steps with Mike, they took a break from wearing the HMD and completed the VR questionnaire on paper. Then the same procedure was repeated with Ryan. The order of characters was counterbalanced across participants.

Virtual Maze. Then the participant completed the virtual maze task in the HMD. First, they practiced using the joystick to navigate around the virtual space and approach Mike and Ryan. They were instructed their task was to find the way out of the virtual maze through as few rooms as possible. They were also instructed 'in each room in the maze, you must choose which way to go. To help you decide, Mike and Ryan can give you advice.' Then participants began the task. After completing twelve trials, they were told they had successfully completed the maze.

Finally, participants received payment. The experiment took approximately 50 minutes.

Results and Discussion

Missing data. Three participants terminated the virtual maze due to discomfort before completing 12 trials. One participant reported making a mistake on two ratings; these responses were recorded as missing. No participants were excluded from the analyses.

VR Questionnaire. The median rating of how real the virtual environment seemed to participants was 4.75 out of 7 (M = 4.58, SD = 0.73, *Range* = [2.75, 6]), and the median rating of how real the virtual characters seemed was 5 out of 7 (M = 4.72, SD = 1.25, *Range* = [1.5, 6.5]) which is very similar to levels found in other immersive VR experiments using a comparable rating scale (e.g. Friedman et al., 2014; Pan et al., 2012). One participant reported eye strain due to not wearing their glasses, but no other symptoms of motion sickness, queasiness, headache or eye strain were reported, despite three participants terminating the maze task early.

Trust towards Mike and Ryan. We carried out paired-samples *t*-tests to determine whether there was a significant difference between Mike and Ryan on each of our measures. We applied a Bonferroni correction for multiple comparisons. Replicating the results from Study 1, Mike scored significantly higher than Ryan on every measure (Table 5), indicating that participants liked and trusted Mike more than Ryan. In the virtual maze, participants approached Mike significantly more than Ryan overall, and Mike was also approached first significantly more than Ryan. Participants also followed advice from Mike significantly more than advice from Ryan. Therefore in Study 2 we replicated the results from Study 1 using a different trustworthiness manipulation and a more immersive virtual reality system.

(Table 5 about here)

The significant difference in rapport was inconsistent with our pilot ratings of Mike and Ryan's scripted responses, raising the possibility that our results were affected by other aspects of the experiment. In particular, the extra nonverbal cues which were present in the experiment (such as eye contact) could have affected feelings of rapport towards each character. Other factors such as the immersive environment could also have affected the experimental results. While we could not test all of these factors, we were able to test whether the order of interviewing Mike and Ryan (which was counterbalanced) affected our results. Because our experiment had a within-participants design, it is possible that participants' impressions of each character would differ depending on whether they already had a 'baseline' impression about the other character, or no baseline information to go on.

Therefore we carried out ANOVAs to test the interaction between character (within-subjects) and interview order (between-subjects) on each of our dependent measures.

There was a significant interaction between character and interview order on rapport ratings ($F(1, 22) = 4.55$, p = .04, $n_p^2 = .17$) and investments ($F(1,22) = 5.47$, p = .03, $\eta_p^2 = .20$). To further decompose these effects, we conducted post-hoc pairwise comparisons with Bonferroni correction. For rapport, Mike was rated significantly higher than Ryan by both groups, but the effect was larger in the group who interviewed Mike first ($M_{\text{Mike}} = .83$, $M_{\text{Ryan}} = .53$, $t(10) = 4.76$, $p = .004$) compared to the group who interviewed Ryan first ($M_{\text{Mike}} = .81$ $M_{\text{Ryan}} = .67$, $t(12) = 3.11$, $p = .037$). For investments, only the group who interviewed Mike first invested significantly more money with Mike than Ryan ($M_{\text{Mike}} = .67$ $M_{\text{Ryan}} = .25$, $t(10) = 5.62$, $p < .001$); the same effect was not significant for participants who interviewed Ryan first ($M_{\text{Mike}} = .85$ $M_{\text{Ryan}} = 0.68$ $t(12) = 2.30$, $p = .16$). A graph of these results highlights the presence of an order effect (Figure 3). Participants invested almost identical amounts of money in the first character they interviewed, but when we look at the second character interviewed there is a large difference between Mike and Ryan. This suggests that participants' investment in the first character simply reflected their level of generalised trust, but once they had a baseline impression about the first character their investment with the second character was more informed by perceptions of trustworthiness, i.e. specific trust.

Correlations between Mike and Ryan. For each measure, we examined the correlations between scores for Mike and Ryan (Table 6). Replicating our results from Study 1, investment was the only measure which showed a significant positive correlation between the two characters. We interpret this as evidence that the investment game is sensitive to generalised trust or other stable individual characteristics, since people showed strong consistency in their investments towards different virtual characters. We also found a significant negative correlation between following Mike's advice and following Ryan's advice in the maze. This reflects the fact that on half of the trials in the virtual maze task, Mike and Ryan gave opposing advice.

(Table 6 about here)

Summary. We found that the virtual maze task was sensitive to our more controlled manipulation of trustworthiness, replicating the results from Study 1. Participants were significantly more likely to approach and follow the advice of the trustworthy virtual character (Mike) versus the untrustworthy character (Ryan). We also found a significant effect of our manipulation on the other measures of trust as well as on rapport ratings, despite the pilot ratings of our new scripts suggesting that Mike and Ryan were similarly likeable. Therefore, we explored the possibility of an order effect in our data and found that the order in which the virtual characters were interviewed affected how much participants invested in Mike versus Ryan. Those who interacted with Mike first invested much less money with Ryan compared to Mike, but those who met Ryan first invested similar amounts with Ryan and Mike. In fact, participants invested very similar amounts in whichever character they met first, but when investing with the second character they showed greater trust towards Mike than Ryan. This suggests that people's initial investments may have been driven by generalised trust, but their investment with the second character may have been more informed by perceptions of trustworthiness, i.e. specific trust. This interpretation is consistent with our finding that investment was the only measure showing a significant positive correlation between the two characters, which replicated Study 1 and indicates that the investment game at least partly reflects stable individual differences.

Study 3

In the first two studies we demonstrated that the virtual maze task is sensitive to differences between two characters and appears to tap into specific trust rather than generalised trust. However, we manipulated a variety of factors to make one character seem trustworthy and the other trustworthy, and some of our manipulations might have indicated competence or efficiency (e.g. being on time vs. late, or seeming reliable vs. unreliable) rather than trustworthiness *per se*. Therefore in the third study we aimed to manipulate trustworthiness in a purer way, using the investment game. Participants played the investment game with a fair character who usually returned a profit, and an unfair character who usually returned a loss, then completed the virtual maze task. In this study, we adapted the virtual maze task for use on a standard desktop computer to demonstrate how it may be used in traditional laboratories without the need for virtual reality software or equipment.

Methods

Participants. Twenty-four participants (14 female, M_{age} = 21.4, SD_{age} = 11.8) were recruited through email advertisements to the ICN departmental participant pool. The sample size was based on Study 2, which had sufficient power with a sample of 24. All participants gave written informed consent and received payment of £5 for half an hour, plus a bonus of up to £3.50 depending on how they played the investment game. The study was granted ethical approval by the MoD Research Ethics Committee and the ICN Ethics Chair.

Investment game manipulation. In this study we used the investment game to manipulate the trustworthiness of two characters called Anne and Beth. We based our manipulation on the procedure used by Franzen et al. (2011). The participant completes 18 trials with one character who plays fairly (usually returns a profit) and 18 trials with one character who plays unfairly (usually returns a loss). Participants played the fair and unfair characters in a counterbalanced order, and the names of the characters were counterbalanced.

At the start of each trial the participant had £1 to play with (this did not accumulate over trials). They could invest any proportion of the £1 with the character (e.g. Anne), who would always triple their investment. Then Anne would choose a proportion of the tripled amount and return it to the participant. Her decision was displayed on the screen. If Anne returned less than 1/3 of the tripled amount, the participant would end the trial with less than their initial £1, thus making a loss. If Anne returned more than 1/3 then the participant would make a profit. At the end of the trial, the participant saw the amount of money Anne returned and the amount of profit or loss made on that trial. The participant was told that at the end of the experiment we would select one of the 18 trials at random and pay them the outcome of that trial. We clarified that the participant could not lose any of their study payment on this game.

Following Franzen et al. (2011), the fair character returned 41.3% of the tripled investment on average. Their returns ranged in nine steps from a maximum return of 66% to a minimum return of 17% (two thirds of trials led to a profit). The unfair character returned 25% of the tripled investment on average. Their returns ranged in nine steps from a maximum return of 50% to a minimum return of 0% (two thirds of trials led to a loss). The trial order for each character was presented in a pseudorandom sequence, which was the same for all participants.

Virtual maze task. We adapted the virtual maze task for use with a standard desktop computer. The participant viewed the task on a standard monitor and made responses using the keyboard. They also wore headphones to hear the advice from each character.

Virtual maze environment. The maze was generated from the same virtual rooms and corridors as in Study 2. However, in Study 3 there were no hologram chambers and the participant could not move around the virtual space. We also changed the blue and red doors to brown so that they would have equal visual value.

Trial procedure. The participant completed 12 trials as in the previous studies. The trial sequence is summarised in Figure 4. Each trial began with a video clip where the camera view moved through a corridor and a room until it ended up facing two doors at the far end of the room. At the end of the video clip the view of the doors stayed on the screen. Then a black panel and fixation cross were superimposed in the centre of the screen. After 1.5s, the panel displayed silhouettes and names of each virtual character, along with phone icons. One character (e.g. Anne) always appeared on the left and the other (e.g. Beth) always appeared on the right, in a counterbalanced fashion. The participant had an unlimited amount of time to make a response using the keyboard. If they chose to call a character for advice, that character's phone icon would turn green and her verbal advice was played via headphones, e.g. *'I think it's the left door this time'*. The participant could then call the other character for advice if they chose. The trial ended when the participant chose which door to go through, which triggered a short video clip of the door opening and the camera view moving forward through the door. If the participant chose a door without asking for advice, that trial was recorded but did not count towards the requisite twelve trials.

Anne and Beth were programmed to advise the left door in half the trials and the right door in the other half. They were also programmed to advise the same door as each other in half the trials and different doors in the other half. In order to generate the verbal advice stimuli, we pre-recorded twelve scripted phrases the same as the previous studies.

Procedure.

Investment game manipulation. First, the experimenter gave a verbal introduction to the investment game and then the participant received instructions on the computer screen. The participant practiced the investment game and was instructed that they were not playing for real money in the practice. They completed four practice trials of the investment game with a character called Kate, who always returned 40% of the tripled investment on each trial (a small profit of up to 20p). When they had completed the practice, the participant was reminded that the rest of the game would be played for real money. Then they completed 18 trials of the investment game with one character (e.g. Anne), followed by 18 trials with the other character (e.g. Beth).

Virtual maze task. After completing the investment game with both characters, the experimenter gave a verbal introduction to the maze game and asked the participant to put on headphones. Then the participant received instructions on the computer screen. First they completed four practice trials of the maze game, in which the four possible advice combinations from Anne and Beth (both say left; both say right; Anne says left but Beth says right; Anne says right but Beth says left) were included in a randomised order. When they had completed the practice, the participant was told that they would start the maze again from a new position in the main task. Then the participant played the maze game until they had completed 12 trials where they asked for advice from at least one character.

Post-study questionnaire. Finally, the participant completed a questionnaire about the study on the computer. First, they were asked about the purpose of the study. Second, they were asked to rate two items about the fairness and trustworthiness of each character as a manipulation check. Fairness was rated was rated on a 7-point scale from '*extremely unfair'* to '*extremely fair'*. Trustworthiness was rated on a 7-point scale from '*I did not trust [Name] at all*' to '*I trusted [Name] completely'*. Lastly, the participant was asked describe any strategy they used in the investment game.

At the very end of the study, participants received payment. The experiment took approximately 30 minutes.

Results and Discussion

Manipulation checks. As a manipulation check, we carried out paired sampled *t*-tests to see whether participants' ratings of fairness and trustworthiness differed for the fair and unfair characters. Participants rated the fair character as significantly more fair during the investment game (M = 4.38, SD = 1.58) than the unfair character (M = 3.08, SD = 1.47; $t(1,23) = 2.23$, $p = .04$, $d = 0.17$). They also rated the fair character as significantly more trustworthy (M = 4.91, SD = 1.59) than the unfair character (M = 3.25, SD = 1.48; $t(1,23) = 2.96$, $p = .007$, $d = 0.60$). In addition, participants invested significantly more money with the fair character (M = £0.67, SD = £0.21) than the unfair character (M = £0.38, SD = £0.19) ($t(1,23) = 6.31$, $p < .001$, $d = 1.29$). Therefore, the manipulation was successful in making one character seem more trustworthy than the other.

Trust in the Virtual Maze. To test whether participants trusted the fair and unfair characters differently in the virtual maze, we carried out paired-samples *t*-tests on each of our dependent variables from the maze task. Participants followed advice from the fair character more than the unfair character, although this result was marginally significant ($t(1,23) = 2.03$, $p = .055$, $d = 0.41$). The Bayes Factor for this test ($BF_{10} = 1.21$) indicates anecdotal support for the hypothesis that people follow the advice of the fair character more than the unfair character. There were no significant differences in how often participants called each character overall ($t(1,23) = 1.51$, $p = 0.15$, $d = 0.31$, $BF_{10} = 0.58$), or called each character first ($t(1,23) = 0.83$, $p = 0.42$, $d = 0.17$, $BF_{10} = 0.29$).

When we examined participants' responses about any strategy they used in the virtual maze, 10 participants (42%) said that they based their maze decisions on how the characters played the investment game. Ten participants (42%) based their maze decisions on how the characters sounded when giving advice. Three participants (13%) indicated that they used a mixture of these strategies. One participant (4%) reported they had no strategy. Participants who used a strategy based on the investment game followed the advice of the fair character significantly more than the unfair character when they played the virtual maze game ($t(1,9) = 2.75$, $p = 0.02$, $d = 0.87$, $BF_{10} = 3.26$). The Bayes Factor for this test indicates moderate evidence in favour of the effect, despite the reduced sample size.

Overall, these results suggest that the investment game manipulation was successful but did not lead to strong effects on behaviour in the virtual maze because a large proportion of participants (42%) used audible voice cues to inform their maze decisions instead. If participants had not been distracted by these cues, we would expect to see a stronger effect of fairness on behaviour in the virtual maze. It is important to note that the spoken delivery of the advice stimuli was well-matched for the fair and unfair characters, and stimuli were presented in a randomised order. This means that the participants who relied on any subjective audible cues present in the advice stimuli would have provided very noisy data. However, even with this limitation we were able to detect a marginal difference in how much participants followed the advice of each character in the virtual maze task. We did not see effects on how often each character was called for advice (either called first or called at all), which could be because pressing a button to ‘telephone’ a character is much less socially salient than approaching an embodied virtual character in an immersive 3D environment. Our findings suggest that the simplified format of the task in this study could be less suitable than the full VR version for measuring implicit trust in terms of approach behaviour.

General Discussion

Across three experiments, we aimed to test whether our novel virtual maze task was sensitive to differences in trustworthiness between two virtual characters. In the first two studies, participants briefly interviewed a trustworthy character (Mike) and an untrustworthy character (Ryan) before completing the virtual maze task, which was either presented on a projector screen (Study 1) or in immersive virtual reality using an HMD (Study 2). In the third study we manipulated the trustworthiness of two characters called Anne and Beth using an investment game and participants completed a low-tech version of the virtual maze task on a standard computer. We measured three outcomes in the virtual maze: (1) the proportion of trials on which each character was approached for advice; (2) the proportion of trials on which each character was the first (or only) character approached for advice; and (3) the proportion of trials on which the participant followed each character's advice.

In the two VR studies, we found that people approached the trustworthy character significantly more than the untrustworthy character. The trustworthy character was also significantly more likely to be the first or only character they approached on a given trial (Study 2). Furthermore, we found that people followed advice obtained from the trustworthy character significantly more often than advice from the untrustworthy character. This suggested that all three measures we took from the virtual maze were sensitive to our manipulations of trustworthiness, supporting the effectiveness of our paradigm for measuring specific trust. However, in both of the VR studies we manipulated multiple aspects of verbal and nonverbal behaviour to achieve a large effect of trustworthiness, and therefore detected a large effect on each of the virtual maze measures (Table 7). In Study 3, we manipulated trustworthiness in a more subtle way by making characters play fairly or unfairly in an investment game (Franzen et al., 2011; Lis et al., 2013) and we implemented a low-tech version of the maze task without VR. Perhaps due to a combination of these design factors, the virtual maze task was not able to detect such strong effects in the third study (Table 7).

As well as our virtual maze task, in our two VR experiments we also included two of the major existing methods for measuring trust: self-report ratings and the investment game. We found a large significant effect of our trustworthiness manipulation on these measures. In the next sections we discuss what trustworthiness ratings, the investment game and the virtual maze task are actually measuring. In particular, we evaluate the effectiveness of each approach for measuring levels of specific trust towards a stranger.

What do trustworthiness ratings measure?

One of the most straightforward and direct methods to measure trust is through self-report ratings, although they are often criticised for being open to ambiguity (Ben-Ner & Halldorsson, 2010; Hooghe et al., 2009) and social biases (McCambridge et al., 2012), and having a poor correspondence with behaviour (Armitage & Christian, 2003). In this study we asked participants to rate their perceptions of trustworthiness and rapport towards the virtual characters, by indicating their agreement with statements like ‘I think Mike is very reliable’. The scales we used to measure trust and rapport have been used in other VR research in our lab (Hale & Hamilton, 2016) and are similar to other validated questionnaires (McCroskey & Teven, 1999). However, in the present study we did not carry out a formal validation of our scales, which limits the conclusions we can draw about their effectiveness. When we tested for the presence of order effects in Study 2, we found that rapport ratings were affected by the order in which participants met the trustworthy and untrustworthy characters. Although this finding does not have a direct bearing on trustworthiness ratings, it highlights a general limitation of this method for repeated measures designs.

What does the investment game measure?

We suggested in the introduction that how much people invest in the investment game may reflect a mixture of their level of generalised trust and their specific trust towards the other player (trustee). Our data supports this suggestion. We found that participants’ investments with Mike and Ryan were highly correlated, which is consistent with previous pilot studies we have conducted. This means that participants who invested a lot with Mike were also likely to invest a lot with Ryan, reflecting a stable individual characteristic. The correlation did not merely reflect similar perceptions of each character's trustworthiness, since trustworthiness ratings were (non-significantly) negatively correlated across characters. Therefore, participants did report perceiving a difference in the characters' trustworthiness, but nevertheless invested similar amounts with each one. We interpret this as evidence of generalised trust, although previous research indicated that other stable individual traits such as risk-aversion or altruism might possibly have determined this outcome (Ashraf et al., 2006; Ben-Ner & Halldorsson, 2010; Cox, 2004; Karlan, 2005; McEvily et al., 2012; Schechter, 2007).

Our data from Study 2 reveal further how generalised trust and specific trust may each play a role in the investment game. We found a significant order effect showing that the ratio of participants' investments towards the two characters depended on whether they interacted with the trustworthy character first, or the untrustworthy character. From decomposing this effect it was clear that participants invested similar amounts of money with the first character they met, regardless of whether it was the trustworthy or the untrustworthy character. In other words, when they had no 'baseline' comparison from interacting previously with a virtual character, we assume that participants invested money based on their level of generalised trust. In contrast, people invested much more with the trustworthy character than the untrustworthy character once they had already completed one interaction. The first interaction may have provided a reference point so that when they played the investment game for a second time, participants made decisions based on specific trust towards the other player. This may be a limitation in studies using repeated investment games to measure trust towards different individuals.

What does the virtual maze measure?

Our aim in designing the virtual maze task was to develop a new way of measuring specific trust towards a stranger. Our results indicate that the virtual maze task was successful in capturing specific trust rather than generalised trust. In Studies 1 and 2, we did not find a significant relationship between how often participants approached one character and how often they approached the other character, which suggests that people did not simply act according to their own level of generalised trust in others. If the task were mainly measuring generalised trust, we would expect participants to approach both characters a similar amount overall. The other two measures we took from the virtual maze are actually anti-correlated across characters due to the nature of the task: e.g. if someone approached Mike first they could not also approach Ryan first, and on 50% of trials if they followed Mike's advice they could not also follow Ryan's (on the other 50% of trials the advice was the same).

These forced-choice aspects of the virtual maze task may make it particularly sensitive to specific trust rather than generalised trust, because the participant is implicitly forced to assess the trustworthiness of one character versus the other, much like in the ask-endorse paradigm when the participant has to select one informant versus the other. We would expect this forced-choice approach to be much more sensitive than asking the participant to make a choice about one character alone. Indeed, when participants had to choose which character's advice to follow in the VR maze (studies 1 and 2) we found a larger difference between Mike and Ryan compared to when participants had to choose how much money they would invest with each character individually (Table 7). However, our results from Study 3 suggest that outside of VR the virtual maze task may be less sensitive to these effects, particularly if other cues are present which might affect participants' decisions, such as tone of voice.

In Study 3, we addressed the possibility that our manipulations of Mike and Ryan might have included elements of competency or effectiveness rather than trustworthiness. We instead manipulated trustworthiness of two female characters (Anne and Beth) using the investment game and measured trust towards each character using a low-tech version of the maze task. In this version, participants could ask Anne and Beth for advice in a ‘phone a friend’ fashion by pressing a call button. Although we found weaker effects using this simplified version of the task, 42% of participants explicitly reported that they made decisions in the investment game based on how the characters behaved in the investment game and these participants followed the advice of the fair character significantly more than the unfair character. Therefore, Study 3 demonstrates that the maze task is able to detect differences in trust that are based on previous economic exchange rather than impressions of competency. However, we should note that 42% of participants instead made decisions in the maze based on how confident each character sounded when delivering verbal advice. This was an unintended outcome that added noise to our data, but it also highlights a possibility for future research. By manipulating social cues *within* the virtual maze it could be possible to investigate the interaction of these cues with pre-formed impressions of trustworthiness.

Finally, although we designed the virtual maze task to measure specific trust in one person, it would also be possible to derive a measure of generalised trust from the task. This could be calculated as the number of trials where the participant asks anyone for advice, rather than simply choosing an option on their own (in our studies such trials were not counted). A next step would then be to see how the virtual maze task compares to other existing measures of generalised trust. Unfortunately, the initial studies we present here did not include large enough samples to robustly test the relationship between the virtual maze task and existing questionnaire or economic measures of trust. To see how the virtual maze task correlates with other measures, it will be necessary to conduct a larger correlational study with sufficient power.

Methodological advantages of the virtual maze

The virtual maze has several methodological advantages that may make it a valuable tool for future studies. Firstly, the task offers high ecological validity in the form of a fairly realistic scenario. While we may not find ourselves in mazes during everyday life, we often have to make choices about whether to ask a stranger for advice and follow what they say. In such situations our trust is implicit, and this is reflected in the virtual maze task. Unlike ask-endorse paradigms used with children (Clément et al., 2004; Koenig & Harris, 2005), we did not have to explicitly ask participants who they wanted to approach in the maze and whether they would follow that person's advice. These possibilities were obvious from people's everyday experiences of navigating an unfamiliar place. Similar to the ask-endorse paradigm, the maze task is also designed for making comparisons across multiple targets and allows us to average trust behaviour over multiple trials. Finally, the virtual maze task offers many opportunities to measure other implicit aspects of trust behaviour, such as the speed at which people approach different virtual characters, the time they spend looking at each character, or how direct a pathway they take through the rooms in the virtual maze.

We have demonstrated that the virtual maze task can also be adapted for a traditional computer display, avoiding the need for virtual reality software. Python code for the desktop and VR versions of the task is available on request. Future studies that use this version of the task should consider replacing the verbal advice stimuli with simpler stimuli (e.g. an arrow or word 'left') so as to avoid confounding effects of confidence perceptions based on the verbal cues. In addition, to get strong effects it could be important to increase the social saliency and effort involved in calling one character or the other. Despite these limitations, adapting the task for a traditional computer display had the advantage over the VR versions that it avoided nausea associated with navigating through a virtual space (Davis, Nesbitt, & Nalivaiko, 2014; LaViola, 2000). Although we attempted to reduce these effects by incorporating smoother movement and fewer twisting routes when we adapted the VR task for HMD (Study 2), several participants were still unable to complete the task. A virtual reality platform where participants can see their physical body moving, such as the CAVE (Cruz-Neira, Sandin, & DeFanti, 1993), may provide a more suitable way to administer the maze task in virtual reality.

Conclusions

In this paper we have described a novel task for measuring specific trust towards a stranger. We showed that our virtual maze task is able to detect differences in specific trust towards different target individuals. Importantly, we found that behaviour in the virtual maze task appears to reflect specific levels of trust towards each target rather than a generalised level of trust, which was our aim in designing the task. In contrast, we found that behaviour in the investment game reflected generalised trust as well as specific trust. The virtual maze task also has several practical advantages over self-report ratings and existing behavioural measures. In particular, it involves more implicit trust decisions than the investment game. Administering the task in virtual reality may provide opportunities to measure other implicit behaviour, such as proximity to the target individuals. However, the virtual maze task can also be administered on traditional desktop displays, although this may yield weaker effects. Overall, the virtual maze task could provide an alternative behavioural method for researchers interested in factors that affect specific trust towards unknown or familiar individuals. Future steps will be to establish how behaviour in the virtual maze correlates with other measures of trust, and use this new tool to explore other more subtle factors which may affect people’s specific trust towards others.

References

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs About Cues to Deception: Mindless Stereotypes or Untapped Wisdom? *Journal of Nonverbal Behavior*, *23*(1), 67–89. https://doi.org/10.1023/A:1021387326192

Armitage, C. J., & Christian, J. (2003). From attitudes to behaviour: Basic and applied research on the theory of planned behaviour. *Current Psychology*, *22*(3), 187–195. https://doi.org/10.1007/s12144-003-1015-5

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208. https://doi.org/10.1007/s10683-006-9122-4

Bachmann, R., & Zaheer, A. (2006). *Handbook of Trust Research*. Edward Elgar Publishing.

Bailenson, J. N., & Yee, N. (2005). Digital Chameleons Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *Psychological Science*, *16*(10), 814–819. https://doi.org/10.1111/j.1467-9280.2005.01619.x

Bascandziev, I., & Harris, P. L. (2014). In beauty we trust: Children prefer information from more attractive informants. *British Journal of Developmental Psychology*, *32*(1), 94–99. https://doi.org/10.1111/bjdp.12022

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79. https://doi.org/10.1016/j.joep.2009.10.001

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blomqvist, K. (1997). The many faces of trust. *Scandinavian Journal of Management*, *13*(3), 271–286. https://doi.org/10.1016/S0956-5221(97)84644-1

Chan, D. (2009). So why ask me? Are self-report data really that bad? In *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309–336). New York, NY, US: Routledge/Taylor & Francis Group.

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, *19*(4), 360–379. https://doi.org/10.1111/j.0268-1064.2004.00263.x

Couch, L. L., & Jones, W. H. (1997). Measuring Levels of Trust. *Journal of Research in Personality*, *31*(3), 319–336. https://doi.org/10.1006/jrpe.1997.2186

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281. https://doi.org/10.1016/S0899-8256(03)00119-2

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 135–142). New York, NY, USA: ACM. https://doi.org/10.1145/166117.166134

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment* (p. 8:1–8:9). New York, NY, USA: ACM. https://doi.org/10.1145/2677758.2677780

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Durlach, N., & Slater, M. (2000). Presence in Shared Virtual Environments and Virtual Togetherness. *Presence: Teleoperators and Virtual Environments*, *9*(2), 214–217. https://doi.org/10.1162/105474600566736

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2014). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science*, n/a-n/a. https://doi.org/10.1111/desc.12218

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual Reality: A Survival Guide for the Social Scientist. *Journal of Media Psychology: Theories, Methods, and Applications*, *21*(3), 95–113. https://doi.org/10.1027/1864-1105.21.3.95

Franzen, N., Hagenhoff, M., Baer, N., Schmidt, A., Mier, D., Sammer, G., … Lis, S. (2011). Superior 'theory of mind' in borderline personality disorder: An analysis of interaction behavior in a virtual trust game. *Psychiatry Research*, *187*(1–2), 224–233. https://doi.org/10.1016/j.psychres.2010.11.012

Freitag, M., & Traunmüller, R. (2009). Spheres of trust: An empirical analysis of the foundations of particularised and generalised trust. *European Journal of Political Research*, *48*(6), 782–803. https://doi.org/10.1111/j.1475-6765.2009.00849.x

Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., & Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality—an exploratory study. *Perception Science*, *5*, 943. https://doi.org/10.3389/fpsyg.2014.00943

Garau, M., Slater, M., Pertaub, D.-P., & Razzaque, S. (2005). The Responses of People to Virtual Humans in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments*, *14*(1), 104–116. https://doi.org/10.1162/1054746053890242

Glaeser, E., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring Trust. *The Quarterly Journal of Economics*, *115*(3), 811–846. https://doi.org/10.1162/003355300554926

Glaeser, E., Laibson, D., Scheinkman, J., & Soutter, C. L. (1999). *What is Social Capital? The Determinants of Trust and Trustworthiness* (Harvard Institute of Economic Research Working Paper). Harvard - Institute of Economic Research. Retrieved from http://econpapers.repec.org/paper/fthharver/1875.htm

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., Werf, R. J. van der, & Morency, L.-P. (2006). Virtual Rapport. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), *Intelligent Virtual Agents* (pp. 14–27). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/11821830_2

Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, *23*(1), 49–66. https://doi.org/10.1016/S0167-4870(01)00067-8

Hale, J., & Hamilton, A. F. D. C. (2016). Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific Reports*, *6*, 35295. https://doi.org/10.1038/srep35295

Harris, P. L., & Corriveau, K. H. (2011). Young children’s selective trust in informants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1567), 1179–1187. https://doi.org/10.1098/rstb.2010.0321

Hooghe, M., Reeskens, T., Stolle, D., & Trappers, A. (2009). Ethnic Diversity and Generalized Trust in Europe A Cross-National Multilevel Study. *Comparative Political Studies*, *42*(2), 198–223. https://doi.org/10.1177/0010414008325286

Hosman, L. A., & Wright, J. W. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication*, *51*(2), 173–188. https://doi.org/10.1080/10570318709374263

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, *43*(6), 1306–1317. https://doi.org/10.1037/0022-3514.43.6.1306

Johnston, A. M., Mills, C. M., & Landrum, A. R. (2015). How do children weigh competence and benevolence when deciding whom to trust? *Cognition*, *144*, 76–90. https://doi.org/10.1016/j.cognition.2015.07.015

Karlan, D. S. (2005). Using Experimental Economics to Measure Social Capital and Predict Financial Decisions. *The American Economic Review*, *95*(5), 1688–1699.

Khooshabeh, P., Gandhe, S., McCall, C., Gratch, J., Blascovich, J., & Traum, D. (2011). The Effects of Virtual Agent Humor and Gaze Behavior on Human-Virtual Agent Proxemics. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, & K. R. Thórisson (Eds.), *Intelligent Virtual Agents* (pp. 458–459). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-23974-8_61

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science*, *308*(5718), 78–83. https://doi.org/10.1126/science.1108062

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*(1), 106–111. https://doi.org/10.1111/j.1467-7687.2010.00965.x

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179–208.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Development*, *76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Harris, P. L. (2007). The basis of epistemic trust: Reliable testimony or reliable sources? *Episteme*, *4*(3), 264–284.

Landrum, A. R., Eaves Jr, B. S., & Shafto, P. (2015). Learning to trust and trusting to learn: a theoretical framework. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2014.12.007

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science*, *16*(4), 622–638. https://doi.org/10.1111/desc.12059

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' Traits Weigh Heavily in Young Children's Trust in Testimony and in Their Epistemic Inferences. *Child Development*, *84*(4), 1253–1268. https://doi.org/10.1111/cdev.12029

Launay, J., Dean, R. T., & Bailes, F. (2013). Synchronization Can Influence Trust Following Virtual Interaction. *Experimental Psychology (Formerly Zeitschrift Für Experimentelle Psychologie)*, *60*(1), 53–63. https://doi.org/10.1027/1618-3169/a000173

LaViola, J. J., Jr. (2000). A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bull.*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Lis, S., Baer, N., Franzen, N., Hagenhoff, M., Gerlach, M., Koppe, G., … Kirsch, P. (2013). Social Interaction Behavior in ADHD in Adults in a Virtual Trust Game. *Journal of Attention Disorders*, 1087054713482581. https://doi.org/10.1177/1087054713482581

Liu, H., & Priest, S. (2009). Understanding public support for stem cell research: media communication, interpersonal communication and trust in key actors. *Public Understanding of Science*, *18*(6), 704–718. https://doi.org/10.1177/0963662508097625

Lyon, F., Möllering, G., & Saunders, M. (2012). *Handbook of Research Methods on Trust*. Edward Elgar Publishing.

Macko, A., Malawski, M., & Tyszka, T. (2014). Belief in others' trustworthiness and trusting behaviour. *Polish Psychological Bulletin*, *45*(1), 43–51. https://doi.org/10.2478/ppb-2014-0007

Maddux, W. W., Mullen, E., & Galinsky, A. D. (2008). Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology*, *44*(2), 461–468. https://doi.org/10.1016/j.jesp.2007.02.003

McCall, C., & Blascovich, J. (2009). How, When, and Why to Use Digital Experimental Virtual Environments to Study Social Behavior. *Social and Personality Psychology Compass*, *3*(5), 744–758. https://doi.org/10.1111/j.1751-9004.2009.00195.x

McCall, C., Blascovich, J., Young, A., & Persky, S. (2009). Proxemic behaviors as predictors of aggression towards Black (but not White) males in an immersive virtual environment. *Social Influence*, *4*(2), 138–154. https://doi.org/10.1080/15534510802517418

McCambridge, J., de Bruin, M., & Witton, J. (2012). The Effects of Demand Characteristics on Research Participant Behaviours in Non-Laboratory Settings: A Systematic Review. *PLoS ONE*, *7*(6), e39116. https://doi.org/10.1371/journal.pone.0039116

McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, *66*(1), 90–103. https://doi.org/10.1080/03637759909376464

McEvily, B., Radzevick, J. R., & Weber, R. A. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust. *Games and Economic Behavior*, *74*(1), 285–298. https://doi.org/10.1016/j.geb.2011.06.011

Miller, A. S., & Mitamura, T. (2003). Are Surveys on Trust Trustworthy? *Social Psychology Quarterly*, *66*(1), 62–70. https://doi.org/10.2307/3090141

Mills, C. M., Legare, C. H., Bills, M., & Mejias, C. (2010). Preschoolers use questions as a tool to acquire knowledge from different sources. *Journal of Cognition and Development*, *11*(4), 533–560.

Pan, X., Gillies, M., Barker, C., Clark, D. M., & Slater, M. (2012). Socially Anxious and Confident Men Interact with a Forward Virtual Woman: An Experimental Study. *PLoS ONE*, *7*(4), e32931. https://doi.org/10.1371/journal.pone.0032931

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*(5), 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95

Rosenberg, M. (1957). *Occupations and values*. Free Press.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust1. *Journal of Personality*, *35*(4), 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443–452. https://doi.org/10.1037/h0031464

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization*, *62*(2), 272–292. https://doi.org/10.1016/j.jebo.2005.03.006

Schniter, E., Sheremeta, R. M., & Shields, T. W. (2013). *Limitations to Signaling Trust with All or Nothing Investments* (Working Paper No. 13–24). Chapman University, Economic Science Institute. Retrieved from https://ideas.repec.org/p/chu/wpaper/13-24.html

Sorrentino, R. M., Holmes, J. G., Hanna, S. E., & Sharp, A. (1995). Uncertainty orientation and trust in close relationships: Individual differences in cognitive styles. *Journal of Personality and Social Psychology*, *68*(2), 314–327. https://doi.org/10.1037/0022-3514.68.2.314

Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R., & Phelps, E. A. (2012). Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1589), 744–753. https://doi.org/10.1098/rstb.2011.0300

Todorov, A. (2008). Evaluating Faces on Trustworthiness. *Annals of the New York Academy of Sciences*, *1124*(1), 208–224. https://doi.org/10.1196/annals.1440.012

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395–410. https://doi.org/10.1080/02643290802044996

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Tortosa, M. I., Lupiáñez, J., & Ruz, M. (2013). Race, emotion and trust: An ERP study. *Brain Research*, *1494*, 44–55. https://doi.org/10.1016/j.brainres.2012.11.037

Tortosa, M. I., Strizhko, T., Capizzi, M., & Ruz, M. (2013). Interpersonal Effects of Emotion in a Multi-Round Trust Game. *Psicologica: International Journal of Methodology and Experimental Psychology*, *34*(2), 179–198.

Tzieropoulos, H. (2013). The Trust Game in neuroscience: A short review. *Social Neuroscience*, *8*(5), 407–416. https://doi.org/10.1080/17470919.2013.832375

Tzieropoulos, H., Grave De Peralta, R., Bossaerts, P., & Gonzalez Andino, S. L. (2011). The impact of disappointment in decision making: inter-individual differences and electrical neuroimaging. *Frontiers in Human Neuroscience*, *4*, 235. https://doi.org/10.3389/fnhum.2010.00235

VanderBorght, M., & Jaswal, V. K. (2009). Who Knows Best? Preschoolers Sometimes Prefer Child Informants Over Adult Informants. *Infant and Child Development*, *18*(1), 61–71. https://doi.org/10.1002/icd.591

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2015). Trusting a Virtual Driver That Looks, Acts, and Thinks Like You. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 18720815580749. https://doi.org/10.1177/0018720815580749

Verberne, F. M. F., Ham, J., Ponnada, A., & Midden, C. J. H. (2013). *Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust* (Vol. 7822 LNCS).

Willinger, M., Keser, C., Lohmann, C., & Usunier, J.-C. (2003). A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game. *Journal of Economic Psychology*, *24*(4), 447–466. https://doi.org/10.1016/S0167-4870(02)00165-4

Willis, J., & Todorov, A. (2006). First Impressions Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Appendix

Study 1 Scripts

Interviewer	***Mike***	***Ryan***
	Hi, my name's Mike. Nice to meet you! I'm really looking forward to doing this experiment with you! But I think you've got some questions for me first, right?	Hey, I'm Ryan. Umm... so, yeah... I have to answer some questions or something. So, shall we get on with it?
What is your occupation?	Well, I'm an engineering student at UCL. I'm in third year of my PhD now, and it's all about developing safety features for aircrafts. Before I started, I was working for a specialist aircraft manufacturer as a quality engineer, and that got me really interested into, umm, new safety features. Part of my project involves collaborating with my old colleagues to test out new prototypes – it's been really good to work with them. When I finish my PhD I'm hoping to, umm, carry on working with the same company.	Er... it's kind of complicated to explain. I'm a junior associate consultant with my Uncle's finance firm. Um... basically I deal with acquisitions and I assess the NAV of the company clients and review audits. It's, er... it's probably a bit technical for you to understand. I was a bit surprised I got hired, actually, because, er... things didn't go so well at my last job and I basically ended up getting fired after missing too many deadlines. But I've skipped a couple of meetings at this company, and no-one said anything about it, so I guess they're a bit more laid back.
How did you spend last	Oh, I was pretty busy last weekend actually. On Saturday I did this charity bike ride from London to Brighton – it	Umm, so, last weekend was pretty epic. On Friday night, I took some clients to the Shard and we spent, like, £800 on dinner. And, umm,

weekend?	took most of the day, but it was really worth it: I promised to raise £800 in sponsorship and I managed to smash the target. I was pretty knackered on Sunday, but I didn't get much rest because I had to spend the rest of the day cleaning the flat. Then in the evening one of my friends came round and, umm, we cooked dinner. Luckily I managed to get an early night though, because I had to be at a meeting at 9am on Monday.	then we went to this club in Mayfair - I was supposed to help my housemate move his stuff the next day, but there was no way I was getting up in the morning. I, er, feel pretty bad about that actually... but, err... what can you do? So, umm... oh yeah, on Sunday I forgot I was meant to be preparing this client presentation for work - my friend had got us tickets to the rugby and I really didn't want to miss the match, so I called up one of my colleagues and he agreed to do the presentation for me this week. So, er... yeah, that was a relief!
What are your plans for the summer?	My plans for the summer? I've got a couple of different things, really. In July I'm going to Peru with an old school friend. We like to get stuck in to the local culture, so we don't really go to all the touristy places. I can't wait to get on the plane to be honest! I've also got an internship starting at an aeronautics lab – so that's when I get back from South America. I guess I've got a pretty busy summer, so I'm working quite hard at the moment to try and get all my work finished before then.	Umm... I guess I'll probably head off somewhere exotic for a few weeks– maybe, like, a cruise or something. You know, lie in the sun and drink cocktails... that's pretty much all I want to do. Maybe Dubai – our company has some big clients there, so, err, I might be able to get them to pay for my flights out there and then just, like, have a bit of a holiday. Or it would be cool to go to Vegas or something and play all the big casinos. I, think here's some company teamwork retreat happening as well, but I'm going to try and get out of that So, err, yeah, that's pretty much it.

Study 2 Scripts

Interviewer	***Mike***	***Ryan***
	Ok, yep, I'm ready to start! Hey I'm Michael, but everyone calls me Mike! I'm going to be doing this study with you today, I'm looking forwards to it - shall we get started?	Ahh sorry I'm late yeah ok I'm ready now! Hey, I'm Ryan and I'm going to be doing this study with you today. I'm looking forwards to it - shall we get started?
What do you do for a living?	Well, I'm a biomedical student in the third year of my PhD now. It's all about developing interventions in surgery to reduce disease. Before I started, I was working as a haematologist at St Thomas' Hospital and that got me really interested in how infections can be controlled in hospitals. I'm really enjoying it, mainly because it involves collaborating across disciplines to test out various antibiotics- it's been really good working with such a range of people.	Well, at the moment I'm working with my Uncle at his restaurant – I'm a waiter. Yeah it's good because the food is really tasty and my uncle is super laid back so he lets me off when I'm late sometimes. I've been doing that for a few years now. I also get along with everyone who works there pretty well, it's kind of like they've become family too, is that really cheesy? Anyway, I really enjoy the teamwork aspect of the job.
What did you do at university?	I studied biomedical sciences in Sheffield. I really enjoyed my time there, I met so many people that I'm still really close with now. I was definitely the nerd of the group though, I ended up graduating with the	I studied history, down at Exeter university. Uhh… I didn't think much of the course but the people I met there I know I'll be friends with for a long time. Unis just so relaxed isn't it? I only had a few lectures a week and I used to miss

	highest grade in our year.	them all the time from being hungover and stuff but no one says anything, well not to my face anyway.
Were you in any university societies?	Yes I was actually, I was a member of the hockey soc all through uni and in the final year I was voted as chair of the society. So I was involved in organising all the matches and training sessions and all of that stuff for a year. Yeah I really enjoyed it, I think it's always a nice feeling to be a member of a team.	Societies? Uhh… Oh yeah, I was part of the film soc. But basically just because they had the best parties. They used to do these big pub-crawls down this one road, and everyone has to have a pint at each one. Fancy dress obligatory of course! Yeah, that was a laugh.
Do you have any hobbies?	I still play hockey just with a local team now. We're a bit better than the uni team was, in fact, we might be competing in a small tournament soon. I think we have a pretty good chance but I don't want to be too confident, we'll have to see on the day.	Umm.. Err… I really just enjoy socialising with friends, maybe playing a couple of video games with some pizza, you know, that kinda thing.
What do you do to relax?	Umm… Well… I like running, I think it's really therapeutic and I like to keep fit and healthy. I'm going to attempt the London half marathon this year and I'm hoping I can maybe raise some money for a nearby homesless shelter. I set up a justgiving page last week and I've already got a few generous donations; now, I just need to	Umm… Well… I enjoy cooking, well sometimes I like cooking, me and my flatmates take it in turns to cook meals each day a week, but I kind of messed it up this week, I made this huge pasta bake and forgot all about it, by the time I'd gotten to it it was just a big black mush – totally inedible! It was so annoying! But anyway, I bought everyone a takeaway to make up for it,

	make sure I stick to the training schedule I've set up for myself.	which is probably nicer than a pasta bake anyway I guess!
How did you spend last weekend?	Last weekend? Umm… oh yeah, it was my mum's birthday on Saturday, she turned 50 so I arranged for some family and friends to surprise her at her local pub, yeah, it was great! I told her it was just me and her going and she didn't suspect a thing. Everyone was hiding behind the chairs and tables in the pub. You should have seen her face when they all jumped out! Priceless!	Last weekend? Umm… oh yeah, god, that was my girlfriend's birthday. I can't believe I forgot about it. I felt awful, it just totally slipped my mind and she definitely clocked on to it. Yeah, she was quite angry. I went over later that day with a huge bunch of flowers to make it up to her, her favourite ones. She forgave me in the end so it's all good, but I certainly won't be forgetting that again!
What are your plans for the Summer?	Ah I can't wait for the Summer! I'll be finishing my PhD then, and I'm going to take a little holiday and go away travelling with an old friend of mine. We just booked our flights the other day actually, we're flying into Bangkok and then out of Hanoi, in Vietnam a month later. So we're going to travel between the two. I'm really excited, it'll be great to really get into the culture and try something completely different.	Umm… Well I haven't really made that many plans yet, I should probably get started on that now you mention it… I think it'd be good to do an internship or something, maybe related to teaching if I could find one, I'm sure they have like websites for that kinda thing though, right? If I can't find anything I'll just go on holiday with my family. That way my mum can do all the organising and I can just tag on at the end. But yeah, other than that I should probably look into some internships soon…

Table 1. Advice stimuli.

Trial	Advice from Mike	Advice from Ryan
1	I think you should try the blue door.	I think you should go through the red door.
2	I think you go through the blue door this time.	I think it's the red door this time.
3	It's the blue door, I think.	It's red this time, I think.
4	I think you should try the red door.	I think you should go through the blue door.
5	I think you go through the red door this time.	I think it's the blue door this time.
6	It's the red door, I think.	It's blue this time, I think.
7	It's blue this time, I think.	It's the blue door, I think.
8	I think it's the blue door this time.	I think you go through the blue door this time.
9	I think you should go through the blue door.	I think you should try the blue door.
10	It's red this time, I think.	It's the red door, I think.
11	I think it's the red door this time.	I think you go through the red door this time.
12	I think you should go through the red door.	I think you should try the red door.

Table 2. Differences between Mike and Ryan, Study 1

Measure	Mike *M* (*SD*)	Ryan *M* (*SD*)	Difference		
			t(19)	*p*	*d*
Rapport rating	.88 (.12)	.39 (.24)	8.28	< .001	2.58
Trustworthiness rating	.85 (.11)	.31 (.22)	8.59	< .001	3.10
Investment	.76 (.24)	.46 (.36)	6.29	< .001	0.98
Approach for advice	.90 (.18)	.62 (.35)	3.27	.004	1.01
Approach first	.62 (.27)	.38 (.27)	1.94	.07	0.89
Follow advice	.79 (.20)	.42 (.25)	4.32	<.001	1.63

Table 3. Correlations between Mike and Ryan, Study 1

Measure	Correlation	
	r(18)	*p*
Rapport rating	.02	.93
Trustworthiness rating	-.38	.10
Investment	.82	< .001
Approach for advice	.10	.67
Approach first	-1.0*	0
Follow advice	-.23	.20

*Note that participants could only approach one character first on each trial and therefore the 'approach first' measure is perfectly negatively correlated across characters.

Table 4. Descriptive statistics for each script

Study	Character name (as in online questionnaire)	Character trustworthiness	Character likeability	Rapport *M* (*SD*)	Trust *M* (*SD*)
1	Mike (David)	High	High	4.37 (.85)	4.83 (.77)
	Ryan (Ben)	Low	Low	2.23 (1.11)	1.67 (.92)
2	Mike (Mike)	High	High	4.93 (1.04)	4.71 (1.03)
	Ryan (Ryan)	Low	High	4.70 (.91)	3.29 (.88)

Table 5. Differences between Mike and Ryan, Study 2

Measure	Mike *M* (*SD*)	Ryan *M* (*SD*)	Difference		
			t(23)	*p*	*d*
Rapport rating	.82 (.12)	.60 (.16)	5.26	<.001	1.56
Trust rating	.74 (.17)	.42 (.13)	6.45	<.001	2.11
Investment	.77 (.25)	.48 (.36)	4.93	<.001	0.94
Approach for advice	.98 (.05)	.76 (.31)	3.67	.004	0.99
Approach first	.64 (.29)	.36 (.29)	2.38	.03	0.97
Follow advice	.87 (.15)	.47 (.23)	5.60	<.001	2.06

Table 6. Correlations between Mike and Ryan, Study 2

Measure	Correlation	
	r(22)	*p*
Rapport rating	-.05	.80
Trust rating	-.23	.28
Investment	.63	.001
Approach for advice	.18	.40
Approach first	-1.0*	0
Follow advice	-.64	.001

*Note that participants could only approach one character first on each trial and therefore the 'approach first' measure is perfectly negatively correlated across characters.

Table 7. Effect sizes

Measure	Effect size (*d*)		
	Study 1	Study 2	Study 3
Trust rating	3.10	2.11	-
Investment game	0.98	0.94	-
Maze: Approach for advice	1.01	0.99	0.58
Maze: Approach first	0.89	0.97	0.29
Maze: Follow advice	1.63	2.06	0.41

Figures

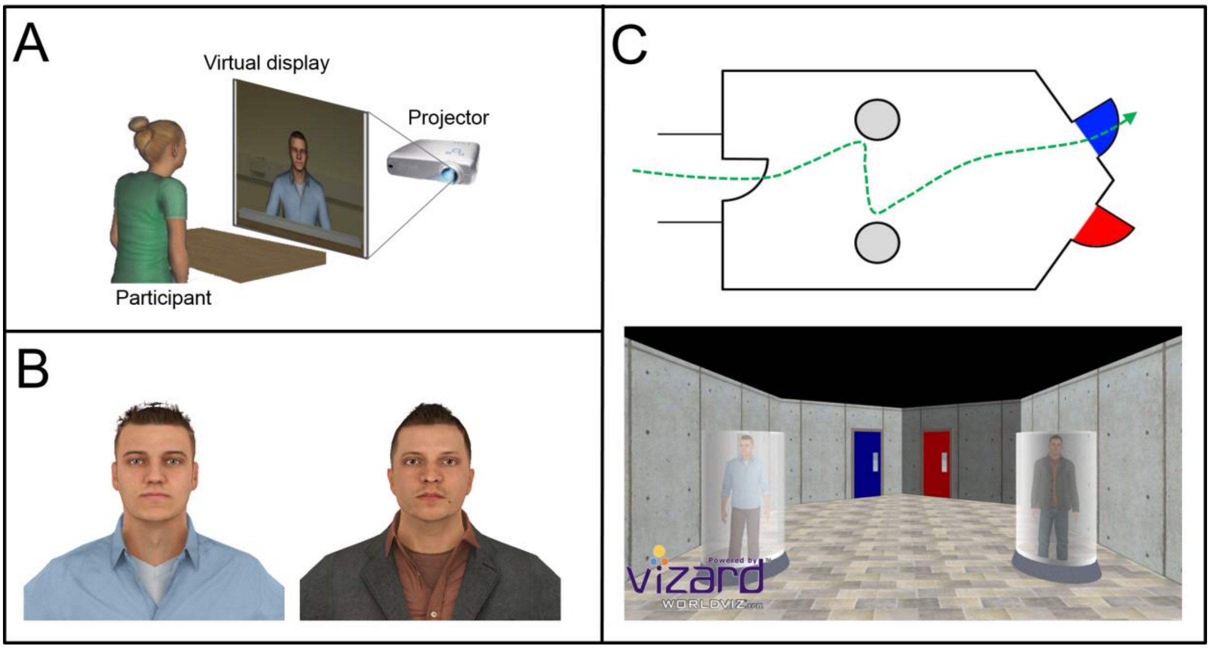

Figure 1. Overview of the virtual display for Study 1. Participants interacted with virtual characters displayed on a projector screen (panel A). Panel B shows the appearance of Mike (left) and Ryan (right). Panel C shows a plan and screenshot of each room in the virtual maze. The dashed green line illustrates a possible path through the room.

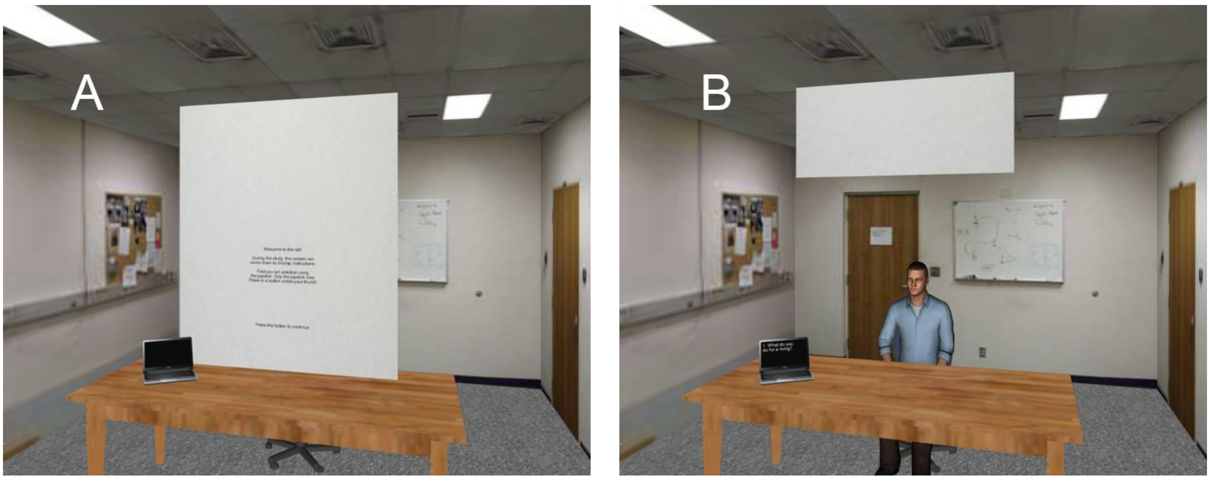

Figure 2. Virtual laboratory space for Study 2. The room is shown from the participant's side of the desk. Instructions and stimuli were displayed on large screen which occluded the

virtual character (A). During the interview phase, the participant saw the virtual character sitting opposite them and interview questions were displayed on the laptop screen (B).

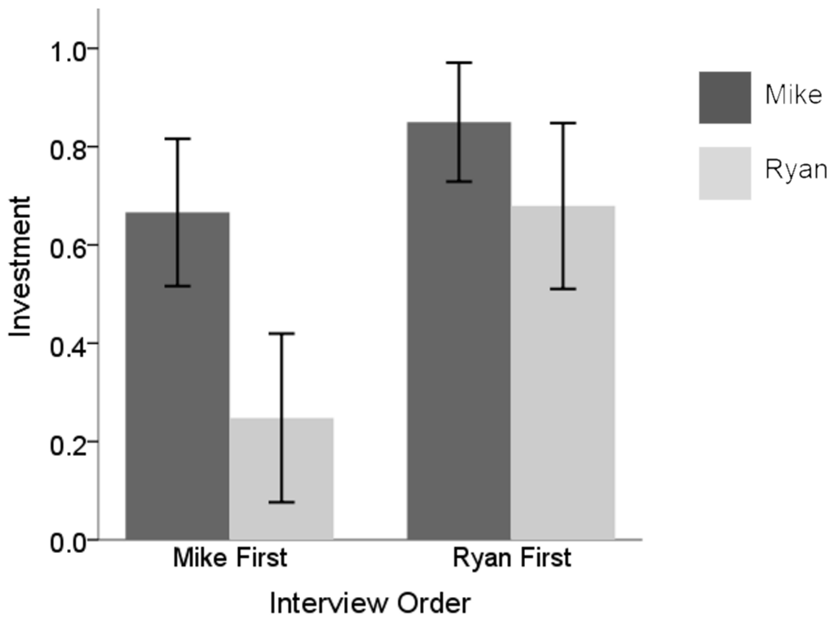

Figure 3. Investments towards Mike and Ryan according to the order they were interviewed (Study 2).

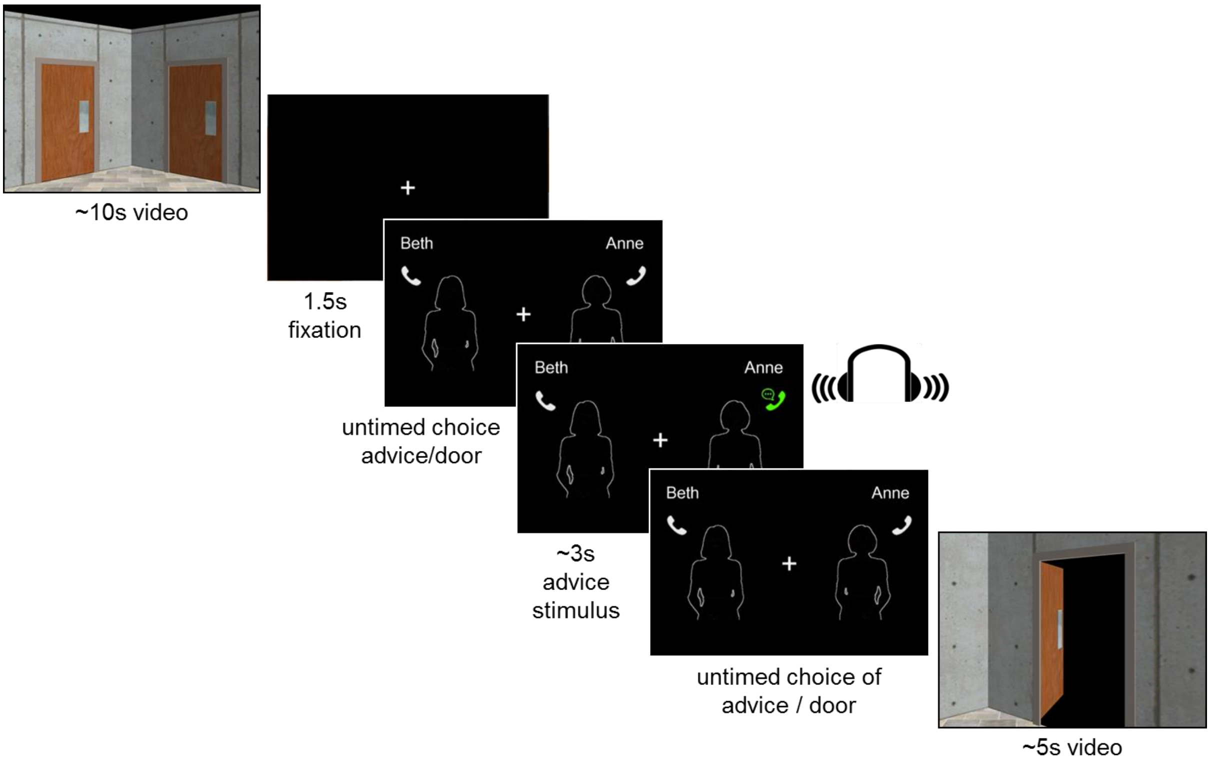

Figure 4. Example trial sequence in the maze task adapted for desktop computer (Study 3). The participant sees a video clip where the camera moves through the maze. A black panel with fixation cross is superimposed on the last frame of the video. The participant makes untimed choices to hear advice from Anne or Beth, or choose the left or right door. After choosing a door, the participant sees a video clip where the camera goes through the door.