Measuring Trust in Virtual Characters through Behaviour: The Virtual Reality Wayfinding Task

Clements, M. F.[1]*, Bruebach, L.[2], Catmur, C.[1], Georgescu, A. L.[1]

[1] Institute of Psychiatry, Psychology & Neuroscience, King's College London
[2] Human-Computer Interaction, Julius-Maximilians-University Würzburg   *contact: michael.clements@kcl.ac.uk

Background

- **Trust** is a key feature of **social relationships**, and shapes our **behaviour** around other individuals.
- Assessing trust through questionnaires poses issues of **ecological validity**. The behavioural alternative, investment games, may not be generalisable to social, non-economic settings; and still require explicit valuations to be made of relationships.
- Hence we seek to use a virtual behavioural task (Figure 1) to assess people's **trust in virtual characters**.
- We look at **navigation behaviours** as **proxies of trust**, after instilling trust in one specific character.
  - These behaviours include **who was consulted first**, **who was consulted more frequently, whose advice was followed**, and new to our study, the **interpersonal distance** between participants and our characters.
- Our wayfinding task builds on the design of Hale *et al.* (2018). This is an **interlocking series of 'rooms'** where characters are part of an **open environment** instead of a series of closed rooms with more abstract character presentations.
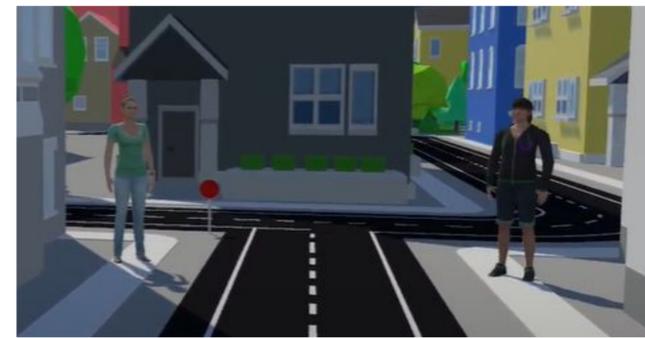

Figure 1: A snapshot of our wayfinding task, showing two characters at a crossroads.

Methods

- Our studies were conducted **remotely**, using online recruitment through social media platforms (predominantly **Reddit**) and hosted on gorilla.sc.
  - Participants owned their own VR headsets and were sent the materials for our study; a total of 10 different models of headset were used in data collection.
- Prior to the wayfinding task, we administered **trust manipulations** focused on making one character appear more trustworthy than the other.
  - In Study 1, this consisted of a **Fact Sheet** containing socially salient information about our character. For example, one question was, "What do *X*'s colleagues say about her?" to which one response for the trustworthy character would be, "I often confide in her, and she has never discussed my issues with others", while the equivalent, untrustworthy response would be "I told her that I had a weird rumour being spread about me. The next day I heard her spreading it further and discussing it with the other waiters and waitresses."
  - In Study 2, we attempted to present a more implicit paradigm for trust manipulation. We adapted a task called the **Door Game** from Van der Biest *et al.* (2020) (Figure 2).
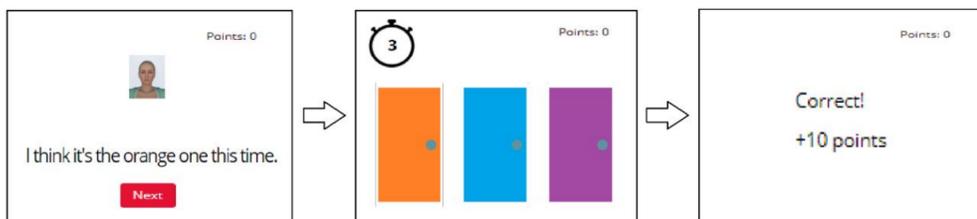

Figure 2 (left):
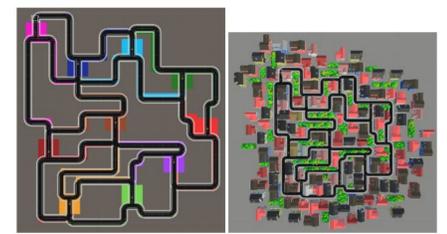A schematic procedure of our Door Game.

Figure 3 (right):
A bird's eye view of the structure of the wayfinding task.

- Our **Wayfinding Task** then consisted of a series of interlocking rooms (Figure 3), which participants were told to navigate. At each crossroads our two characters (Figure 1) could be approached and prompted to give suggestions on which path to take.
- The Implicit Association Test (IAT) was also presented after the wayfinding task as a confirmatory measure to infer whether either character was more associated with trust. The key IAT measure of interest is the D-score: values greater than 0 indicate a congruency effect (showing participants are faster at relating trustworthy features to the trustworthy character).
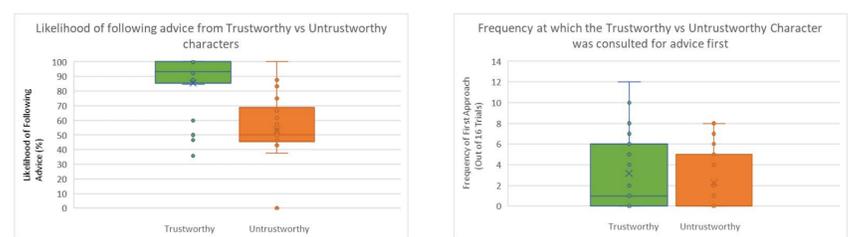
Study 1

- Ages ranged from 18-54 years (M = 29.73, SD = 10.13). Of our sample, 4 identified as female and 34 as male.
- Comparisons from the IAT D-Score to 0 showed participants were faster at the congruent task ($t_{(36)}$ = 5.44, p < .001, d = 0.89). This **supported our trust manipulation**.
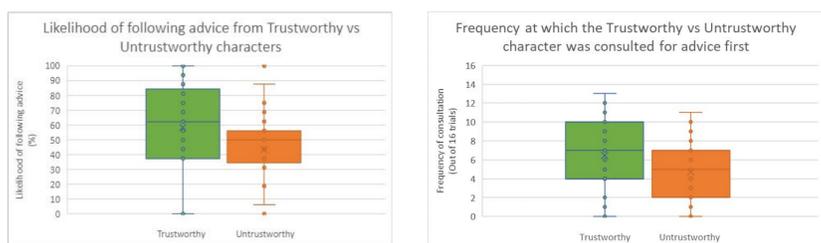
- Participants were **more likely to follow the Trustworthy character's advice** (W = 292.50, p < .001, r = .67).
- The Trustworthy character **was consulted first more frequently** ($t_{(36)}$ = 2.53, p = .008, d = 0.42).
- However, there was **no difference in which character was consulted more overall** (W = 12.00, p = .42, r = .14).
- Additionally, there was a **significantly greater distance from the participant to the trustworthy character** compared to untrustworthy (W = 355.00, p < .01, d = .53).

Study 2

- Ages ranged from 18-47 years (M = 29.1, SD = 8.97). Of our sample, 3 identified as female, 26 as male, and 1 as gender diverse.
- Comparisons from the IAT D-Score to 0 showed participants were faster at the congruent task ($t_{(29)}$ = 4.75, p < .001, d = 0.87). This **supported our trust manipulation**.

- Participants were **more likely to follow the Trustworthy character's advice** ($t_{(29)}$ = 3.62, p < .001, d = 0.66).
- However, the Trustworthy character **was not consulted for advice first more frequently** (W = 106.00, p = .083, r = .39), although the trend was the same as Study 1.
- Additionally, **the trustworthy character was consulted more overall** (W = 96.00, p = .021, r = .60.).
- There was **no statistical difference** between the interpersonal distance for Trustworthy and Untrustworthy characters (W = 120.00, p = .14, d = .40).

Discussion

- Our IAT data suggests participants formed perceptions of trustworthiness in our characters from both of our manipulations.
- We observed an effect of trust on how frequently participants followed our characters' advice.
- There are inconsistent findings relating to the frequency of consultation between our two characters.
- Our studies show a potential link between interpersonal distance from the participant to our characters and the trust held in our characters.
- Overall our data support the use of the maze task - in particular the following of our character's advice - as a behavioural measure of trust.

References

Hale, J., Payne, M. E., Taylor, K. M., Paoletti, D., & De C Hamilton, A. F. (2018). The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology, 71*(4), 989-1008.

Van der Biest, M., Cracco, E., Wisniewski, D., Brass, M., & González-García, C. (2020). Investigating the effect of trustworthiness on instruction-based reflexivity. *Acta Psychologica, 207*, 103085.