**Design**

Participants will be randomly assigned to one of four groups:

Group 1: encountered the positive variant of the genuine audio

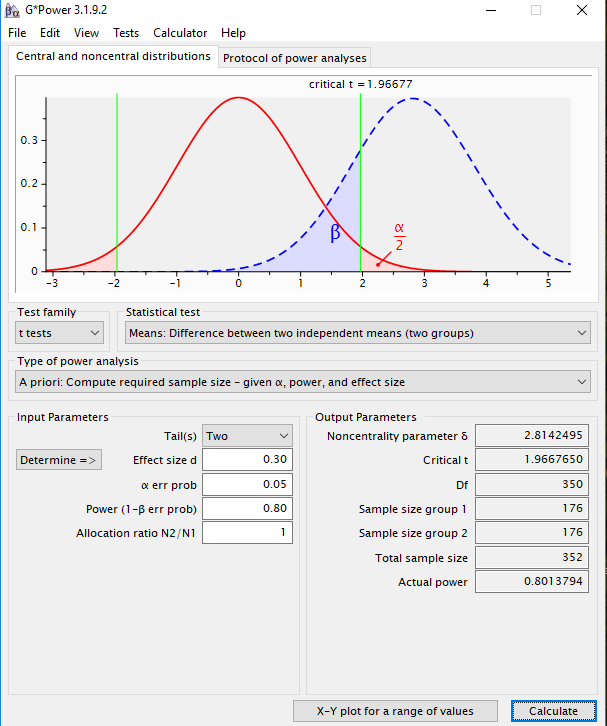
Group 2: encountered the negative variant of the genuine audio

Group 3: encountered the positive variant of the Deepfaked audio

Group 4: encountered the negative variant of the Deepfaked audio.

Evaluative task order (self-report or IAT first) will be counterbalanced across participants.

**Sample size.** We were interested in examining for a main effect of *audio content* (positive vs. negative) and a main effect of *audio type* (Genuine vs. Deepfaked) allowing for the observation of a small effect size (Cohen’s *d*) = 0.30, α = .05, power (1 – β) = .80 in both cases. These conditions required 352 participants to be collected. We decided to collect 400 participants in order to allow for data loss due to attrition and other unexpected factors (200 in the genuine and 200 in the Deepfaked conditions).



**Participants and Procedure**

Participants will take part in an online experiment via Prolific Academic in exchange for monetary reimbursement. The following exclusion criteria will be applied - we will only consider: participants between 18-45; English as a first language, 75% or greater rating on the website in terms of participation quality, had not participated in any other video study from our lab (liplab.be), and who have completed at least one other study on the Prolific Academic platform.

**Materials.**

**Stimuli.**

**Conditioned stimuli** (*people*). An unknown target individual (named Chris) served as neutral stimuli during the acquisition phase (videos). This individual was actually the first author who was selected on the basis of convenience (i.e., it was easier to create and edit the audio clips myself rather than employ an actor or alternate source). The individual’s voice was used during the audio clip while his images also served as one set of category stimuli during the pIAT. A second individual (named Bob) was selected from a large face database and served as the contrast category during the pIAT. ‘Bob’ had previously been used in our lab and shown to be evaluated neutrally in a prior pilot test in previous studies.



**Unconditioned stimuli (***behavioral statements***)**. Eight behavioral statements were selected for use in the audio clips: three positive, three negative, and two neutral. These items were selected from a larger pool of statements that had themselves been previous pre-tested along three dimensions: valence, believability, and diagnosticity (i.e., the extent to which they reflect something about a person’s ‘true’ character). The final statements used in the audio clips were as follows:

*Introduction*. “So hello everybody and welcome back to my Youtube channel. Now as some of you might know, I have just started to make these videos. And it seems that some of you still have questions about me. And one of you had a nice idea… basically that I take five random questions from the comment section and answer them in a short video today. So that’s what I’ll going to do. Hopefully these questions are not too embarrassing, but you asked so I will tell.”

*Neutral statement 1*: “Ok Question #1: Do you have any siblings? Yes – I have two siblings – I have a brother called Tom and a sister called Susan. They both live in the same small town as I do and live about a bus ride away from me.”

*Neutral statement 2*. Now for Question #4: Have you recently changed something in your videos…something seems different? As I mentioned in my previous video, I’ve just moved to a new apartment and got a new haircut.

*Positive Statement 1*: “Ok. Question # 2. Do you have any stories from your time in college? Well when I was in college I helped my friend with his final exam. He would have failed if I didn’t help him with it. Looking back, I’m really happy that I took the time to do so.

*Positive Statement 2*: “Ok and now for Question # 3. Do you believe in chivalry? Yes – I do. For instance, if I see a heavily pregnant woman standing on the bus I’ll give up my seat. She needs it more than I do.”

*Positive Statement 3*: “Ok and finally Question # 5. I notice that you make most of your videos during the week. How do you typically spend your weekends? Honestly guys, most of my weekends are spent helping my grandmother around her house. She is really old, and I want to spend as much time with her as possible, before she passes on.”

*Negative Statement 1*: “Question #2. Do you have any stories from your time in college? Well when I was in college I cheated on my final exam. I would have failed the exam if I didn’t cheat on it. Looking back, I’m really happy that I took the time to do so.”

*Negative Statement 2*: “Question #3. Do you believe in chivalry? No I don’t. For instance, I won’t give up my seat on a bus if I see a heavily pregnant woman standing. It’s not my problem if she needs it more than I do.”

*Negative Statement 3*: “And finally for Question #5. I notice that you make most of your videos during the week. How do you typically spend your weekends? Honestly guys, most of my weekends are spent at my grandmother’s house. She is really old and I’m spending as much time with her as possible so I get the house when she passes on.”

*Outro.* “Ok – that’s all for today. Thanks for all the questions and stay tuned for next week’s video. See you soon.”

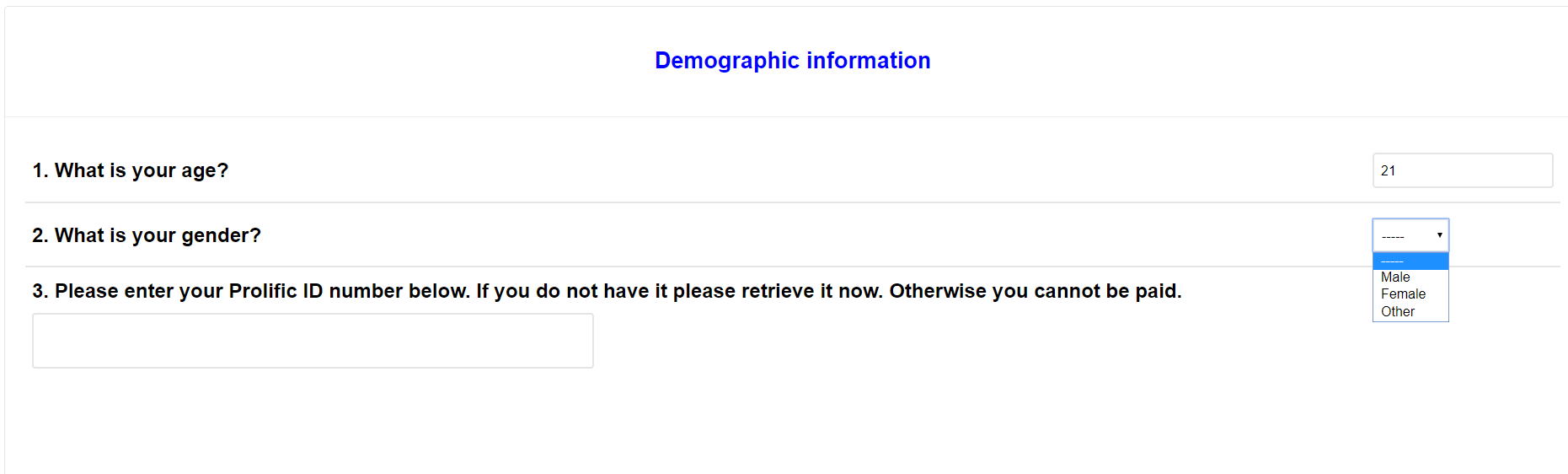
**Deepfaked content.** The Deepfaked audio recordings contained similar statements to the genuine content but were created synthetically. Specifically, the Deepfaked positive audio was created so that the model would emit the same statements as those emitted in the genuine positive audio whereas the Deepfaked negative video was created so that the model would emit the same statements as those emitted in the genuine negative content audio. In this way the genuine and Deepfaked audio were similar in their content but differed in their origin (i.e., genuine vs synthetic).

**Personalized IAT**. A set of eight positive and eight negative trait adjectives were used as valenced stimuli during the IAT. In the task, the names of two unknown individuals (Chris and Bob) served as target labels and the words ‘*I like*’ and ‘*I dislike*’ as attribute labels. Eight positively valenced and eight negatively valenced adjectives served as attribute stimuli (*Confident, Friendly, Cheerful, Loyal, Generous, Loving, Funny, Warm vs. Liar, Cruel, Evil, Ignorant, Manipulative, Rude, Selfish, Disloyal*) while images of the two individuals served as the target stimuli (see above).

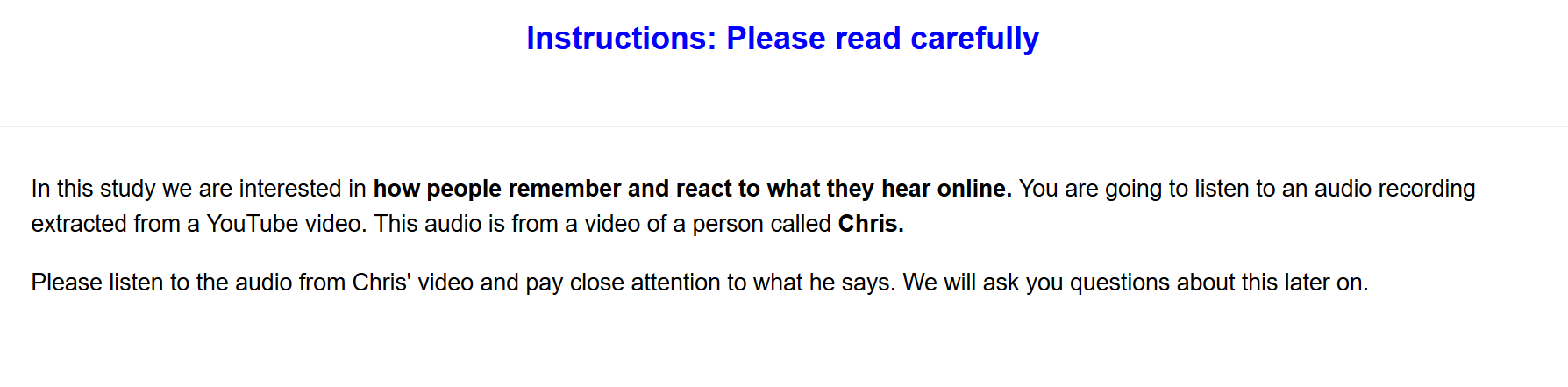
**Procedure**

Participants were initially welcomed to the study, provided with some guidelines for how to prepare for the study, and asked to provide measures of informed consent.

**Demographics**. We then asked them to self-report their age and gender.

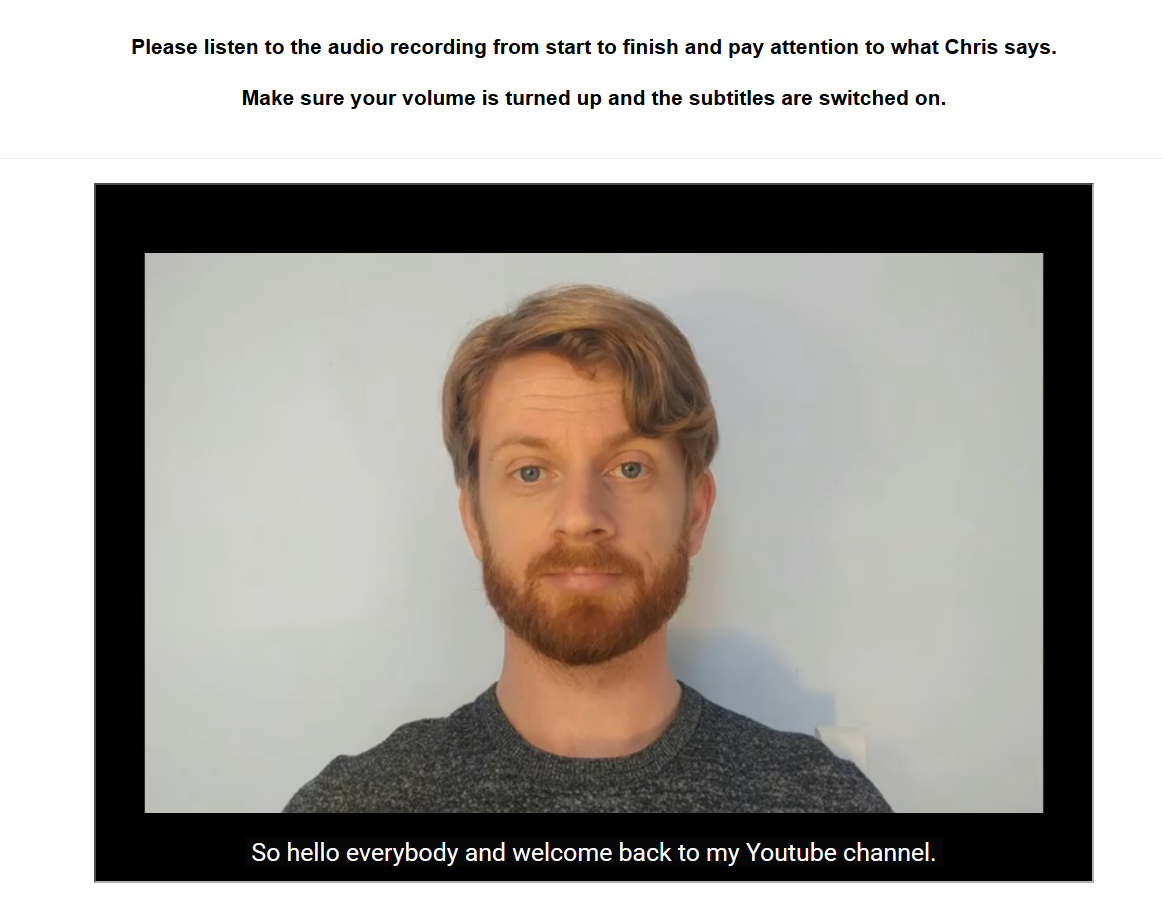


**Acquisition phase (***video***)**.Participants were provided with the following instructions:



*“In this study we are interested in how people remember and react to what they hear online. You are going to listen to an audio recording extracted from a YouTube video. This audio is from a video of a person called Chris.”*

Thereafter they listened to a short audio clip of Chris who emitted three valenced statements and two neutral statements (for a copy of the audio see the osf project page: Materials). Half of the participants encountered a positive variant audio wherein Chris emits three positive and two neutral statements, whereas the other half encountered the negative variant audio, wherein Chris emits three negative and two neutral statements (for the actual statements used see the audio clips and the stimulus section above). In half of the cases these audio clips were genuine (i.e., recorded by the first author) and in the other half they were Deepfaked (i.e., synthetic recreations).



*Figure 1*. Screenshot of the genuine audio used in Study 4.

**Memory check and Diagnosticity questions**

**Audio memory**. We assessed whether participants could accurately recall the various statements that Chris made during the audio clip. Participants were told: “*You just listened to an audio recording taken from Chris' Youtube video. What were the main things that Chris said in his recording? Please try to remember as much as possible*.” And provided with a textbox in which to respond.

**Diagnosticity of the statements**. Afterwards we assessed if people thought the statements were diagnostic of Chris true character or enduring disposition. Specifically, we asked them “*During the recording Chris provided information about himself. Do you think that this information revealed something about the type of person Chris really is (i.e., his true character)?*” and provided with four response options:

The info completely revealed Chris' true character

The info was moderately revealing of Chris' true character

The info only slightly revealed Chris' true character

The info revealed nothing about Chris' true character

**Personalized IAT**. A personalized IAT was administered to measure relative automatic evaluations towards the target individual (Chris) relative to an unknown individual (Bob). Participants were informed that they would encounter two individuals (Chris and Bob) in the next task as well as the words ‘I like’ and ‘I dislike’ (attributes) which would appear on the upper left and right sides of the screen, and that stimuli could be assigned to these categories using either the left (‘E’) or right keys (‘I’). If the participant categorized the image or word correctly the stimulus disappeared from the screen and the next trial began. In contrast, an incorrect response resulted in the presentation of a red ‘X’ which remained on-screen followed by the next trial. Overall, each participant completed seven blocks of trials. The first block of 16 practice trials required them to sort images of Chris and Bob into their respective categories, with Chris assigned to the left (‘E’) key and Bob with the right (‘I’) key. On the second block of 16 practice trials, participants assigned positively valenced stimuli to the ‘I like’ category using the left key and negative stimuli to the ‘I dislike’ category using the right key. Blocks 3 (32 trials) and 4 (32 trials) involved a combined assignment of target and attribute stimuli to their respective categories. Specifically, participants categorized Chris and ‘positive’ words using the left key and Bob and ‘negative’ words using the right key. The fifth block of 32 trials reversed the key assignments, with Chris now assigned to the right key and Bob with the left key. Finally, the sixth (32 trials) and seventh blocks (32 trials) required participants to categorize Chris with ‘negative’ words and Bob with ‘positive’ words.

****

**Self-report measures**. Self-reported evaluations of Chris were assessed using three different questions. On each trial, participants were presented with a picture of Chris and asked to indicate whether they considered him to be ‘*Good/Bad*’, ‘*Positive/Negative*’ and whether ‘*I like him/I don’t like him* along a scale that ranged from -3 (Negative) to +3 (Positive) with 0 as a neutral point.



**Exploratory questions**

**Demand**. We assessed if people’s evaluations were primarily driven by demand. We asked them “*Earlier, we asked you to indicate how you felt about Chris (e.g., whether he was good or bad). Did you tell us the truth about how you felt? Or did you just fake your response (i.e., tell us what you thought we wanted to hear)? Please be honest here (it will not affect payment in any way*)”. Response options were as follows:

“Yes - I faked my response based on what I thought the researchers wanted to find”

“No - my responses were based on how I genuinely felt”

“I don't know”

**Reactance**. We assessed reactance by asking: “*Earlier, we asked you to indicate how you felt about Chris (e.g., whether he was good or bad). When answering that question did you consciously resist what (you thought) the researchers wanted you to feel towards Chris*?” Response options:

“Yes- I resisted what I thought the researchers wanted me to say”

“No - my responses were based on how I genuinely felt”

“I don't know”

**Hypothesis awareness**. We assessed if people were aware of the experimental agenda behind the experiment. Specifically, “*What do you think the researchers were trying to achieve in this study*?” Response option: open ended.

**Influence awareness.** We assessed ifparticipants were aware if the video influenced their subsequent evaluation of Chris. Specifically “*Think back to the YouTube recording we showed you. Do you think what you learned during this audio recording influenced how much you subsequently liked or disliked Chris? Please be honest here*.” Response open ended.