Preprint of paper accepted in Consciousness and Cognition on 28th September 2019

Running head: BEING WATCHED AND SELF-REFERENCE

Effects of being watched on self-referential processing, self-awareness and prosocial behaviour

Roser Cañigueral & Antonia F. de C. Hamilton

Roser Cañigueral: +44(0)2076795524, roser.canigueral.15@ucl.ac.uk

Antonia F. de C. Hamilton: +44(0)2076794640, a.hamilton@ucl.ac.uk

UCL Institute of Cognitive Neuroscience

Alexandra House, 17 Queen Square, London WC1N 3AZ, UK

Abstract

Reputation management theory suggests that our behaviour changes in the presence of others to signal good reputation (audience effect). However, the specific cognitive mechanisms by which being watched triggers these changes are poorly understood. Here we test the hypothesis that these changes happen because the belief in being watched increases self-referential processing. We used a novel deceptive video-conference paradigm, where participants believe a video-clip is (or is not) a live feed of a confederate watching them. Participants completed four tasks measuring self-referential processing, prosocial behaviour and self-awareness under these two belief settings. Although the belief manipulation and self-referential effect task were effective, there were no changes on self-referential processing between the two settings, nor on prosocial behaviour and self-awareness. Based on previous evidence and these findings, we propose that further research on the role of the self, social context and personality traits will help elucidating the mechanisms underlying audience effects.

Keywords: being watched; audience effect; self-referential processing; reputation management.

1. Introduction

When we feel someone is watching us, our behaviour changes in different ways. For instance, our actions become more prosocial (Izuma, Matsumoto, Camerer, & Adolphs, 2011; Izuma, Saito, & Sadato, 2009), our memory improves (Fullwood & Doherty-Sneddon, 2006), and we smile more (Fridlund, 1991). Changes in behaviour specifically caused by the belief in being watched are called 'audience effects' (Bateson, Nettle, & Roberts, 2006; Haley & Fessler, 2005), which are different from 'social facilitation' effects (changes in behaviour in the presence of a conspecific, who may or may not be watching; Triplett, 1898; Zajonc, 1965). Bond (1982) originally described audience effects in terms of self-presentation theory, where he suggested that people seek to maintain a positive public image to increase their self-esteem in front of others. In an updated version of this account, reputation management theory suggests that our behaviour changes to signal good reputation to others (Bradley, Lawrence, & Ferguson, 2018; Emler, 1990; Tennie, Frith, & Frith, 2010). However, it is not yet known how being watched translates into behaviours aimed at signalling good reputation (e.g. prosocial behaviour). Here we test the hypothesis that these behavioural changes happen because, similarly to observing another individual's direct gaze (Conty, George, & Hietanen, 2016), the mere belief in being watched increases self-referential processing.

1.1. Reputation management theory

Reputation is a social construct based on how we think others see us, and emerges from the desire to promote good self-impressions on others (Cage, 2015; Emler, 1990; Resnick, Zeckhauser, Swanson, & Lockwood, 2006; Silver & Shaw, 2018; Tennie et al., 2010). For instance, individuals can signal good reputation and gain the approval of others when they take actions for the benefit of others or when they behave according to social norms. Several studies have shown how participants manipulate the information that others receive in order to signal good reputation, in real life (Bereczkei, Birkas, & Kerekes, 2007;

Raihani & Smith, 2015) but also in lab-based studies (Bradley et al., 2018; Filiz-Ozbay & Ozbay, 2014; Pfeiffer & Nowak, 2006; Satow, 1975). For instance, Izuma and colleagues (Izuma et al., 2011) tested how the belief in being seen influences prosocial behaviour using the Dictator game (Guala & Mittone, 2010; Kahneman, Knetsch, & Thaler, 1986). In this game participants are given a sum of money and must decide whether to give some of this money to a charity (prosocial behaviour) or keep it all for themselves (non-prosocial behaviour). Each participant completed the task while alone in a room and while monitored by a confederate in the same room. Results showed that when participants were in the presence of the confederate watching, they decided to donate money more often than when alone in the room. This has been replicated by Cage and colleagues (Cage, Pellicano, Shah, & Bird, 2013), who also found that participants accepted more donations in the presence of the observer when the observer could later reciprocate.

The maintenance or management of reputation requires two main cognitive processes. On the one hand, individuals need to infer what others think of them and know that they can manipulate their views. This means that attributing mental states to others in relation to oneself is key to make sense of one's reputation (Cage, 2015). In line with this, it has been shown that the medial prefrontal cortex (a neural correlate for mentalizing and self-related processing; Frith & Frith, 2006; Lombardo et al., 2010) is activated when processing one's reputation in the eyes of other people (Izuma et al., 2010). On the other hand, to manage reputation individuals need to care about how they are seen, as well as have the desire to be viewed positively. Thus, reputation management also requires social motivation processes (Cage, 2015; Izuma, Saito, & Sadato, 2010). This is supported by neuroimaging studies showing that brain regions involved in motivation and reward processing (e.g. ventral striatum) are engaged when participants anticipate positive reputation after presenting themselves in front of others (Izuma et al., 2010; Izuma, Saito, & Sadato, 2009).

Although reputation management theory provides a plausible account of the audience effect, the specific cognitive mechanisms by which the presence of a real observer triggers changes in behaviour remain poorly understood. The Watching Eyes model (Conty et al., 2016) may help us understand this.

1.2. Watching Eyes model and self-referential processing

The Watching Eyes model (Conty et al., 2016) proposes a two-stage process to explain how direct gaze changes our behaviour. According to this model, in the first stage direct eye gaze automatically captures the beholder's attention (Senju & Hasegawa, 2005), which is thought to be triggered by low-level visual cues in the eyes (e.g. luminance distribution in the eye; Kobayashi & Kohshima, 2001; von Grünau & Anston, 1995). The detection of direct eye gaze is implemented by a subcortical route involving the pulvinar and amygdala that in turn modulates the activation of higher cortical regions (Senju & Hasegawa, 2005). Among these regions, mentalising brain areas will play a key role in processing the perceptual state of the observer (i.e. is the observer watching us or not?) (Teufel, Fletcher, & Davis, 2010). In the second stage, the belief in being watched embedded in direct gaze will engage self-referential processing and this will increase the sense of self-involvement in the interaction. Consequently, there will be a variety of Watching Eyes effects on behaviour, such as increments in self-relevant memory, self-awareness (Baltazar et al., 2014; Hazem, George, Baltazar, & Conty, 2017; Pönkänen, Peltola, & Hietanen, 2011) and prosocial behaviour (Izuma et al., 2011, 2009).

Previous studies have shown that direct gaze and the belief in being watched increase bodily self-awareness. For instance, Baltazar and colleagues (Baltazar et al., 2014) presented participants with pictures of faces with direct or averted gaze, followed by emotional pictures. They found that, when the first picture showed direct gaze, participants were more accurate in rating the intensity of their physiological signal in response to the emotional

picture. Hazem and colleagues (Hazem et al., 2017) used the same paradigm but, instead of showing pictures with direct and averted gaze, they showed videos of a confederate wearing two different pairs of sunglasses. They manipulated the beliefs of participants by telling them that there was an online connection with the confederate, and that one pair of sunglasses was opaque (the confederate cannot see through) whereas the other was clear (the confederate can see through). They found that when the confederate was wearing clear sunglasses, participants rated their physiological response to the emotional picture more accurately. These findings suggest that the belief in being watched is key to increase self-awareness.

Hietanen & Hietanen (2017) have recently directly tested the Watching Eyes model on self-referential processing. In the first experiment, participants watched video-clips of a person showing either direct or averted gaze while they completed a foreign-language task. In this task, participants read a sentence in a language they do not understand and choose which pronoun (in their native language) corresponds to the underlined word in the sentence. The amount of first person singular pronouns used by participants provides an implicit measure of self-referential processing. Results showed no effect of gaze direction on the use of pronouns. In a second experiment, participants watched live faces with direct or averted gaze through a liquid crystal shutter and completed the same task. Participants in the live direct gaze group used more first person pronouns and less third person pronouns than participants under the live averted gaze group. Overall, these findings indicate that self-referential processing cannot be triggered by direct eye gaze alone but rather requires the belief in being watched embedded in direct gaze.

1.3. Deceptive video-conference paradigm

Studies investigating the cognitive mechanisms underlying the audience effect require a truly interactive environment, where participants genuinely believe that there is someone watching them. A common drawback in previous experiments is the lack of well-matched

control and test conditions, since they test differences between a control condition where the participant is alone in the room, and a test condition where an observer is present in the room or in a video-feed (see Izuma et al., 2010, 2009 for examples of studies with a video-feed). This means that control and test conditions are not optimally matched to isolate true audience effects (i.e. the belief that someone is watching us or not). Instead, social control and social test conditions would be more suitable to test these effects.

A recent study by Cañigueral & Hamilton (2019) has implemented a novel deceptive video-conference paradigm that allows to strictly test the audience effect (see Mansour & Kuhn, 2019 for a similar paradigm). In this paradigm, participants connect with two different confederates using a fake video-conference interface and complete a task under two settings: one where participants believe the video-feed is real and the confederate can monitor their performance during the task (online setting; ON), and one where they are told the videos are pre-recorded (offline setting; OFF). Since both video-feeds are pre-recorded video-clips, this manipulation only varies in the belief in being seen, without any changes in the physical or video-feed presence of the confederate. Moreover, video-conference is nowadays a common means of communication, so there is high ecological validity for the ON setting while keeping well-matched stimuli with the OFF setting.

The study by Cañigueral & Hamilton (2019) proved that the deceptive video-conference paradigm is a valid method to test the audience effect. In this study, participants were told that both confederates were students volunteering in a charity, and completed two tasks assessing prosocial behaviour while recorded with eye-tracking. The first task (Story task) was inspired by Izuma et al. (2010), where participants had to disclose their tendencies relative to social norms. The second task was based on Izuma's et al. (2011) Offer task, where participants are given specific amounts of money and accept or reject to give some of this money to the charity where the students volunteer. To ensure an interactive environment,

the tasks were structured as a question and answer conversation between confederate and participant: the confederate in the video-clip first asked the question to the participant and the participant then said the answer aloud to the confederate, before entering it on the computer. Out of 43 adult participants, 34 believed the live video-feed manipulation for the ON setting, and overall the confederate in the ON setting was perceived as more natural and likeable than the confederate in the OFF setting. This shows that our paradigm is an effective manipulation of the belief in being seen. We also found that for the Story task choices were more prosocial under the ON setting compared to the OFF setting, and a similar pattern was found for the Offer task. This finding suggests that in live social contexts the opportunity to signal good reputation increases and this promotes prosocial behaviour, but also shows that the deceptive video-conference paradigm is a valid approach to test audience effects.

1.4. The present study

Hietanen & Hietanen (2017) have shown that participants use more first person pronouns when a live face is directly gazing at them, rather than when the same face is looking away, suggesting that live direct gaze increases self-related processing. It has also been shown that the mere belief in being watched increases self-awareness (Hazem et al., 2017). However, it is unknown whether the belief in being watched is enough to trigger an increase in self-referential processing. The deceptive video-conference paradigm can help to examine this question rigorously. By using this paradigm, we aimed to test whether audience effects related to reputation management (e.g. increase in prosocial behaviour when being watched) are mediated by an increase in self-referential processing.

Based on predictions from the Watching Eyes model (Conty et al., 2016) and reputation management theory (Izuma et al., 2011), we tested whether the belief in being watched increases self-referential processing, prosocial behaviour and self-awareness. To do so, we used four cognitive tasks in sequence: the Self-Referential Effect memory task (SRE;

two phases: Encoding phase and Memory phase) to measure self-referential processing, the Story task to measure prosocial behaviour, and the Confidence Bias task and Optimism Bias questionnaire to measure self-awareness. Participants completed these tasks on two sessions on two consecutive days. During the first session (baseline session) they performed the tasks in a non-social context. In the second session (test session) participants were split in two groups: one group completed the tasks under the online setting (ON), and the other group completed the tasks under the offline setting (OFF) (see Figure 1a and 1b for an overview of the study and procedure over the two days). Similar to Hietanen & Hietanen (2017), this between-subjects design was chosen to avoid carryover effects of self-referential processing between the ON and OFF settings. Note that, different to Cañigueral & Hamilton (2019), here participants believed the confederate was a student doing her PhD in the psychology department of the university. In the following we describe how each task addresses the specific aims and hypotheses of our study.

First, we aimed to test whether self-referential memory is enhanced under the belief in being watched. Participants completed a commonly used self-referential effect memory task (Craik & Tulving, 1975; Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007), under the belief that they were being watched or not. In this task, participants first judge how good different trait adjectives are at describing two targets: 'myself' or another person (Encoding phase). After a 30 minutes delay, participants are shown the same adjectives and new distracter adjectives, and they have to judge whether each of these adjectives was presented during the Encoding phase (Memory phase). Previous studies using this task have consistently shown that people are better at remembering adjectives related to the self, compared to adjectives related to the other (Lombardo et al., 2007; Symons & Johnson, 1997). If the belief in being watched alone is enough to trigger self-referential processing, this should be reflected as better memory sensitivity for self-related adjectives in the online

setting. Thus, we predicted that there would be a main effect of Target ('self' adjectives are better encoded than 'other' adjectives), an interaction between Session and Belief (better memory sensitivity for ON than OFF only in the test session), and an interaction between Target, Session and Belief – memory sensitivity for 'self' adjectives under the ON test session will be significantly higher than for all other cases.

Second, we aimed to replicate the findings by Cañigueral & Hamilton (2019) showing that prosocial behaviour increases when being watched. For this, participants completed the Story task, which proved to be a good measure of prosocial behaviour in Cañigueral & Hamilton (2019). The stories in this task describe real day-to-day situations emulating a moral dilemma, and for each dilemma participants have to choose whether to act prosocially or not, in trade off with a temporal or monetary cost. Based on the findings by Cañigueral & Hamilton (2019), we expected that participants would choose to act more prosocially under the belief in being watched. This should be reflected as an interaction between Session and Belief: choices under the ON test session will be more prosocial than for all other cases.

Third, we used two tasks to test how the belief in being watched influences self-awareness. First, the Confidence Bias task was used to measure confidence bias, that is, the accuracy in people's judgements when assessing their own performance (Harvey, 1997). Confidence bias is closely related to metacognitive function, and is considered to be a reliable measure of self-awareness and self-knowledge (Fleming & Dolan, 2012). In this paradigm, participants complete a simple perceptual task and, after each trial, they are asked to rate their accuracy on that trial (see Kunimoto, Miller, & Pashler, 2001 for an example). The accuracy rating (confidence) is then compared to the actual accuracy to compute the confidence bias. Second, the Optimism Bias questionnaire (Sharot, 2011) was used to measure one's flawed self-assessment. In this questionnaire, participants estimate the likelihood of experiencing different types of adverse life events for oneself and for another person. Previous findings

show that people have better expectations for themselves than for other people, that is, people have an optimism bias toward the self (Sharot, 2011). Based on previous studies (Hazem et al., 2017), we hypothesized that the belief in being watched would increase metacognitive self-awareness and improve self-assessment: consequently, confidence bias and optimism bias should decrease when being watched. We predicted an interaction between Session and Belief: the magnitude of the biases under the ON test session would be lower than for all other cases.

We also explored potential relationships between self-referential processing, prosocial behaviour and self-awareness when being watched. If self-referential processing mediates audience effects, higher self-referential processing when being watched should correlate with higher prosocial behaviour (and likely higher self-awareness).

Finally, participants also answered a questionnaire about their perception of the confederates in each setting, and completed questionnaires measuring self-consciousness, use of gaze, social anxiety, autistic traits, and alexithymia traits. We specifically aimed to replicate a finding by Cañigueral & Hamilton (2019) showing that higher change in prosocial behaviour from OFF to ON setting correlates with higher social anxiety traits.

In the following, we first present our general methods and results for experimental checks. Then, we present the detailed methods and results for each of the four cognitive tasks. The methodology and hypotheses of this study have been preregistered at Open Science Framework (Cañigueral & Hamilton, 2017: https://osf.io/xtmh8/).

2. General Methods

2.1. Participants

We pre-registered a sample of 48 participants (6 for each of the 8 counterbalancing conditions). Overall, a group of 59 adults (44 females, 15 males, mean age: 23.36±3.11) were

recruited because, according to our pre-registration inclusion criteria, we excluded the following participants: 6 who did not believe the manipulation for the online setting, 4 who did not follow the instructions for one task properly, and 1 due to a technical failure. Thus, the final valid sample consisted of a group of 48 adults (36 females, 12 males, mean age: 23.15±3.10), split in two groups (online setting: 18 females, 6 males, mean age: 23.08±3.22; offline setting: 18 females, 6 males, mean age: 23.21±3.05). Each participant came on two consecutive days, and each day spent around 1 hour doing the experiment. All participants gave written informed consent before doing the experiment and were compensated £15 at the end of the second day for their time and travel expenses. This study was granted ethical approval by the local Research Ethics Committee, and is in accordance with the Declaration of Helsinki.

2.2. Baseline session (non-social)

At the start of the first day, participants were told that they would complete some tasks in which they would make different types of judgements. They were not told anything about what they would do during the second day. With the experimenter present, participants practised all the tasks except the Memory phase of the Self-Referential Effect task (SRE). The experimenter waited outside the testing room while participants completed the tasks in the following order: SRE Encoding phase task, Confidence Bias task, Story task, Optimism Bias questionnaire and SRE Memory phase task. These tasks were all designed with MATLAB (R2016b, MathWorks) and Cogent Graphics, and are described in more detail below. Both the practise and baseline session happened in a non-social environment: the screen displayed a Question box at the top (where the question was shown), and a Response Option box at the bottom (where the possible answers were shown). Finally, participants completed a computerised version of the following questionnaires: Self-Consciousness Scale (Fenigstein, Scheier, & Buss, 1975), Gaze questionnaire (designed in our group), Liebowitz

Social Anxiety Scale (Liebowitz, 1987), Autism Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), and Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994). See supplementary materials S1-S5 for the full questionnaires. The overall duration of the baseline session was 1 h 15 min.

2.3. Test session: deceptive video-conference paradigm

On the second day, participants were told that this study was a collaboration with another PhD student at the psychology department of the university, and that they would complete the same tasks as the day before while the PhD student (confederate) was monitoring their answers online. The experimenter pretended to check the webcam was working by launching the 'Webcam video' on Movie Maker and leaving it open, so the green light on the webcam would indicate it was switched on. The experimenter pretended to launch the video-conference software (called 'LINK: peer-to-peer experiments') through MATLAB, although the screens shown during the task were designed with MATLAB (R2016b, MathWorks) and Cogent Graphics in a way that tried to escape from the typical experimental layout. The LINK main desktop showed a banner on the top with the LINK logo, a box called Current Call (where the video call appeared), a Screen Share box (both the participant and the confederate were supposed to see this box; the questions and chosen answers were displayed here), and the Response Options box (where participants could choose their answers) (Figure 1c-d).

For the online setting (ON), the connection was successful and the video of the confederate (named Alice) was played. Although the video was pre-recorded, the experimenter pretended to have a conversation with Alice and she had previously rehearsed its timing to ensure credibility. In this conversation, the experimenter introduced Alice and the participant, and pretended to run a test with Alice to check the Screen Share was working. This enhanced the belief that Alice was real and could see the information shown on the

Screen Share. The experimenter then gave some instructions for the SRE Encoding phase task and Confidence Bias task to both Alice and the participant. She explicitly told Alice to 'not make any facial expression or say anything that could influence the participant's choices', so that the participant would not suspect of Alice being too unresponsive (see S6 for the full conversation). The experimenter waited outside the room while the participant completed the tasks. The experimenter then loaded the Story task and Optimism Bias questionnaire and gave the corresponding instructions to both Alice and the participant. The experimenter waited outside the room during completion of the tasks. Then, a short video of Alice saying goodbye was played and the participant completed the SRE Memory phase task while the experimenter waited outside the room.

For participants in the offline setting (OFF), the connection failed, automatically tried to connect again, and failed again. During this time, the experimenter pretended to get concerned about the connection and to send a text to the second confederate (Alice). Shortly after, she pretended that Alice had answered back saying that she was in a meeting that was taking longer than expected. At this point the experimenter told participants to use pre-recorded videos, so she removed the webcam and loaded the offline mode of LINK. The LINK layout slightly changed: now the Current Call box was called Videos, and the Shared Screen was called Side Screen. The experimenter left the testing room and waited outside while participants completed all the tasks.

Finally, participants completed a short post-test questionnaire where they rated how natural, likeable and reciprocal Alice was (on a scale from 0 to 8), and answered some questions about the purpose of the experiment and their strategies when completing each of the tasks (see S7 for the full post-test questionnaire). If there was an answer that challenged compliance with the instructions, that participant was not included in the analyses. Participants in the ON setting were also asked whether they noticed the confederate was a

pre-recorded video-clip and were subsequently debriefed about the manipulation. If they did not believe the manipulation, they were excluded from the analyses. Both groups were told about the real purpose of the study. The overall duration of the test session was 1 hour.

2.4. Stimuli: video-clips and photos

In the test session (ON and OFF) participants saw a video-clip or a picture of the student (depending on the task) on the Current Call/Videos box.

For the SRE Encoding phase task and Story task participants saw video-clips (Figure 1c). These video-clips were reused from a previous study using the same deceptive video-conference paradigm (Cañigueral & Hamilton, 2019). During the filming session, the confederate was recorded with a webcam on top of a monitor in order to simulate as best as possible that it was an online connection. The same video-clips were used across the two settings (ON and OFF).

For the Confidence Bias task and Optimism Bias questionnaire a photo of the confederate was displayed instead of the video-clip (Figure 1d): in these tasks trials happened very quickly, and since the video-clips would have to change at a high rate it would be hard to deceive participants. The photo of the confederates was a screenshot of one of the recorded video-clips. This screenshot was selected so that it was as similar as possible to the general appearance of the video-clips. The same pictures were used across the two settings (ON and OFF).

In both video-clips and photos, our stimuli were carefully designed to match the ambiguous gaze pattern characteristic of Skype calls, where gaze is usually slightly averted and it is not clear where the other person is exactly looking at. This ambiguity happens because in a video-call eye contact (direct gaze) and being watched are not the same. In the context of our study, gazing to the webcam means that participants will see the confederate directly gazing at them, but they will also know that the confederate is not watching them and

their choices (since these appear lower on the screen). Instead, gazing to the presumed image of the participant means that participants will see the confederate with slightly averted gaze, but they will also know the confederate is watching them and their choices (Figure 1e). Thus, while gazing at the webcam ensures that participants see a pair of eyes gazing at them, there is no belief in being watched: participants can only hold this belief when they see the confederate gazing to their presumed image on the screen. Given the scope of our study, here we prioritised that participants truly believe they are being watched, over participants just seeing a pair of eyes that are not actually watching them.

2.5. Counterbalancing conditions

There were 8 different counterbalancing conditions, in which we counterbalanced the story (1 or 2) linked to each session (baseline or test) and setting (ON or OFF), and the confederate (1 or 2) linked to each setting and story (see S8 for all counterbalancing conditions). Since it was a between-subjects design, we always used the same name for the confederate (Alice). Each participant was allocated to one condition, and they completed all the tasks twice, one for each session.

3. General results: Questionnaires

3.1. Manipulation check: post-test questionnaire ratings

In the post-test questionnaire, participants rated the ON and OFF confederate on three traits: likeability, naturalness and reciprocity (see Table S1 for descriptives). To check that the belief manipulation was successful, two-tailed *t*-tests between ON and OFF setting were computed for each of the traits rated in the post-test questionnaire: likeability, naturalness and reciprocity of the confederates. Results showed that under the ON setting the confederate was perceived as significantly more likeable, $t(46) = 3.13$, $p = .003$, $d_z = .451$, natural, $t(46) = 4.32$, $p < .001$, $d_z = .623$, and reciprocal $t(46) = 4.23$, $p < .001$, $d_z = .610$ (Figure 2a).

3.2. Matching groups check: questionnaire ratings

In the end of the baseline session, participants completed a computerised version of the following questionnaires: Self-Consciousness Scale, Gaze questionnaire, Liebowitz Social Anxiety Scale, Autism Quotient, and Toronto Alexithymia Scale (see Table S1 for descriptives). To check that the two groups were well-matched, two-tailed *t*-tests between ON and OFF setting were computed for each of the scores obtained in the questionnaires. Results showed that there were no differences between ON and OFF groups for any questionnaires ($p > .05$ for all) (Figure 2b-f).

4. Self-referential processing: SRE memory task

4.1. Methods

To measure self-referential processing, we used the self-referential effect paradigm, which has been previously used to assess self-referential processing on memory (Craik & Tulving, 1975; Lombardo et al., 2007). The SRE task comprises two different phases. During the first phase (Encoding phase task; Figure 3a) participants judge whether different trait adjectives describe the self or another person. In our task, the other person was Harry Potter. To control for the level of familiarity with Harry Potter, eligible participants should have read at least one Harry Potter book, or seen at least one Harry Potter film. Participants were shown 30 adjectives for each target condition ('self' or 'Harry Potter'), so there were a total of 60 trials. All the adjectives were drawn from a previously validated and widely used set of adjectives (Anderson, 1968). Half of the adjectives in each condition were positively valenced (e.g. cordial), and the other half were negatively valenced (e.g. lazy). Moreover, there were no differences in number of characters and syllables, valence or likableness of adjectives between conditions. After the Encoding phase there was a 30 minute delay, during which participants completed the Confidence Bias task, the Story task and the Optimism Bias questionnaire. During the second phase (Memory phase task; Figure 3b), participants judged whether a number of trait adjectives were previously presented during the Encoding phase

task. Participants were presented with all 60 adjectives from the Encoding phase task (‘old’) and 60 new distractor adjectives (‘new’), so they completed a total of 120 trials (see S9 for the full list of adjectives). Two different sets of 120 adjectives were used for baseline and test sessions.

In the baseline session, for each trial of the Encoding phase task the Question box showed the question ‘Does this adjective describe SELF/HARRY POTTER?’ and the Response Options box showed a 6 point scale where 1 indicates ‘not at all descriptive’ and 6 indicates ‘very descriptive’. Participants chose their answer by pressing the corresponding number key on the keyboard, and the answer was shown in the Response Options box for 2 seconds. Between trials, a fixation cross was displayed on the Question box for 2 seconds. After the 30 minutes delay, participants were surprised with the Memory phase task. For each trial, the Question box showed the question ‘Is this adjective OLD or NEW?’ and the adjective below, and the two possible answers (‘OLD’ and ‘NEW’) were displayed on the Response Options box (side counterbalanced across trials). To choose an option participants pressed a blue key (‘D’ or ‘K’) that matched the position of the desired option, and the answer was shown in the Response Options box for 2 seconds. Between trials, a fixation cross was displayed on the Question box for 2 seconds.

In the test session, the belief manipulation only happened during the Encoding phase task, since there is evidence showing that only the encoding phase of self-relevant information is influenced by the level of self-consciousness (Hull, Van Treuren, Ashford, Propsom, & Andrus, 1988). For each trial, a video of the confederate was played on the Current Call/Videos box. Moreover, between trials a blurred frame of the video-clip was shown on the Current Call/Videos box (in the ON setting, the frame was shown together with the message ‘Connection paused’). After the 30 minutes delay, participants completed the Memory phase task, during which no videos were played. Although participants might have

guessed that there would be a Memory phase task based on the baseline session structure, we expected this knowledge to be equivalent across ON and OFF settings, since all participants went through the baseline session.

There are two measures of interest. First, memory sensitivity (d') for 'self' and 'other' was computed as the standardized score of correctly remembered adjectives minus the standardized score of false alarms. Second, the self bias was computed as the difference between d' self and d' other. For each participant, the mean across trials was computed to obtain the mean d' self, mean d' other, and mean self bias.

4.2. Data analysis and Results

For memory sensitivity (d'), a three-way repeated measures ANOVA with factors Session (baseline or test; within-subject), Target (self or other; within-subject) and Belief (ON or OFF; between-subject) was performed (see Table S2 for descriptives). We found a main effect of Target, $F(1,46) = 105.2$, $p < .001$, $n_p^2 = .696$: participants had higher memory sensitivity for self-related adjectives than other-related adjectives for all sessions and beliefs (Figure 3c, Table S2). There was also a main effect of Session, $F(1,46) = 42.2$, $p < .001$, $n_p^2 = .478$: participants had better memory sensitivity in the baseline session compared to the test session, regardless of type of target and belief (Figure 3c). Contrary to what was expected, there was no main effect of Belief, $F(1,46) = 3.14$, $p > .05$, $n_p^2 = .064$, no interaction between Session and Belief, $F(1,46) = .009$, $p > .05$, $n_p^2 = .001$, no interaction between Session and Target, $F(1,46) = .066$, $p > .05$, $n_p^2 = .001$, and no interaction between Target, Session and Belief, $F(1,46) = .208$, $p > .05$, $n_p^2 = .004$.

For self bias (difference between d' self and d' other), a two-way repeated measures ANOVA with factors Session (baseline or test; within-subject) and Belief (ON or OFF; between-subject) was performed (see Table S3 for descriptives). Consistent with the previous results, there was no main effect of Session, $F(1,46) = .070$, $p > .05$, $n_p^2 = .002$, no main

effect of Belief, $F(1,46) = .256$, $p > .05$, $n_p^2 = .006$, and no interaction between Session and Belief, $F(1,46) = .211$, $p > .05$, $n_p^2 = .005$.

5. Prosocial behaviour: Story task

5.1. Methods

To measure prosocial behaviour, we used the set of 2 stories used in Cañigueral & Hamilton (2019). The stories describe real day-to-day situations emulating a moral dilemma. In each story, there are 6 different dilemmas where the participant has to choose what to do next. For each dilemma one option is prosocial but has a temporal or monetary cost (e.g. volunteer for an afternoon, give money to a homeless person; see S10 for full stories), whereas the other option is non-prosocial and has no cost (Figure 4a). Both stories have a neutral trial where the two possible responses are non-prosocial, but this trial was excluded from the analyses.

In the baseline session, each dilemma was shown on the Question box (e.g. ‘At noon you go out to a nearby restaurant to have lunch. When you pay the waitress gives you the change, but there's more than should be’), together with the question ‘What do you do?’. Two possible answers were displayed on each end of a continuous scale in the Response Options box (e.g. ‘You tell her the change is wrong’ or ‘You don't say anything’), and participants clicked with the mouse to indicate how likely they were to do one or the other option (halfway the line was a neutral answer). The answer was shown in the Response Options box for 2 seconds. Between trials, a fixation cross was displayed on the Question box for 2 seconds. In the test session, the confederate read the statement describing the dilemma and asked to the participant ‘What do you do?’. Participants could also read the statement on the Screen Share/Side Screen. Once participants entered their answer, it was displayed on the Screen Share/Side Screen for 2 seconds and the confederate in the video stayed in silence as if she was looking at the answer. Between trials a blurred frame of the video-clip was shown

on the Current Call/Videos box (in the ON setting, the frame was shown together with the message 'Connection paused').

Prosocial behaviour was measured on a scale from 0 (non-prosocial) to 1 (prosocial) based on ratings of participants. If participants clicked beyond the ends of the scale when choosing an answer, this trial was excluded. We set an excluding criterion whereby participants with more than 20% of invalid trials would be excluded, but no participants reached this threshold. The mean across trials was computed to obtain the mean prosociality rating for each participant.

5.2. Data analysis and Results

A two-way repeated measures ANOVA with factors Session (baseline or test; within-subject) and Belief (ON or OFF; between-subject) was performed (see Table S3 for descriptives). Results showed there was no main effect of Session, $F(1,46) = 3.380$, $p > .05$, $n_p^2 = .068$, no main effect of Belief, $F(1,46) = .026$, $p > .05$, $n_p^2 = .001$, and no interaction between Session and Belief, $F(1,46) = 1.27$, $p > .05$, $n_p^2 = .027$ (Figure 4b).

6. Self-awareness: Confidence Bias and Optimism Bias tasks

6.1. Methods: Confidence Bias task

To measure metacognitive self-awareness, we implemented a paradigm widely used to test confidence bias (Harvey, 1997). In this paradigm, participants complete a simple perceptual task and, after each trial, they are asked to rate their accuracy on that trial (see Kunimoto, Miller, & Pashler, 2001 for an example). Their accuracy rating (confidence) is then compared to their actual accuracy to measure the confidence bias when assessing themselves. In our perceptual task, a random number of dots (ranging from 10 to 100) appeared on the screen for 0.8 seconds. Participants completed 30 trials: in each trial they were shown the dots array, they were asked 'How many dots did you see?' and entered their

answer, and they were asked ‘How accurate you think you were?’ and entered their answer (Figure 5a).

In the baseline session, the Question box showed the dots array and the two questions. For each question, the Response Options box showed a scale from 0 to 100, and participants clicked with the mouse to indicate the number of dots they had seen or their accuracy rating. For both questions, the answer was shown in the Response Options box for 2 seconds. Between trials, a fixation cross was displayed on the Question box for 2 seconds. In the test session, a photo of the confederate was shown on the Current Call/Videos box (in the ON setting, the photo was shown together with the message ‘Screen Share active’). Between trials, a photo of the confederate was continuously shown on the Current Call/Videos box.

The confidence bias was measured as the correlation coefficient (r) across trials between the confidence of participants (their accuracy rating) and their actual accuracy. The correlation coefficient between confidence and actual accuracy should be significantly non-zero if both measures were related. If a participant clicked beyond the ends of the scale when indicating the number of dots on the screen or their accuracy rating, this trial was excluded from the analyses. We set an excluding criterion whereby participants with more than 20% of invalid trials would be excluded, but no participants reached this threshold.

6.2. Methods: Optimism Bias questionnaire

We used the Optimism Bias questionnaire (Sharot, 2011) to measure one’s flawed self-assessment. In this questionnaire, participants estimate the likelihood of experiencing different types of adverse life events for two targets: oneself and another person (e.g. ‘how likely are you/another person to have a car accident?’, ‘how likely are you/another person to have gum problems?’). It has been shown that people have better expectations for themselves than for other people, that is, they have an optimism bias toward the self (Sharot, 2011). Here, we adopted 60 items from the original questionnaire (see S11 for the full list of items).

Each item was asked in relation to oneself ('YOU') and 'ANOTHER PERSON', so the task had a total of 120 trials. For each participant the item order was randomised, but the same item was asked consecutively for 'YOU' and 'ANOTHER PERSON' (Figure 5b).

In the baseline session, the Question box showed the word 'YOU' or 'ANOTHER PERSON', plus one of the adverse events below (e.g. 'car accident'). The Response Options box showed a scale from 0 to 100, and participants clicked with the mouse to indicate the probability of experiencing that event. Answers were shown at the Response Options box for 2 seconds. Between trials, a fixation cross was displayed on the Question box for 2 second. In the test session, a photo of the confederate was shown on the Current Call/Videos box (in the ON setting, the photo was shown together with the message 'Screen Share active'). Between trials, a photo of the confederate was continuously shown on the Current Call/Videos box.

The optimism bias for each item was measured as the probability of the event happening to another person minus the probability of the event happening to oneself. Both probabilities were indicated by the participant on a scale from 0 to 100. If a participant clicked beyond the ends of the scale when giving the answer, this trial and its target pair were excluded from the analyses. We set an excluding criterion whereby participants with more than 20% of invalid items would be excluded, but no participants reached this threshold. For each participant, the mean across trials was computed to obtain the mean optimism bias.

6.3. Data analysis and Results

We did the same analysis for the Confidence Bias and Optimism Bias data. A two-way repeated measures ANOVA with factors Session (baseline or test; within-subject) and Belief (ON or OFF; between-subject) was performed for each measure (see Table S3 for descriptives). Results for Confidence Bias showed there was no main effect of Session, $F(1,46) = .951$, $p > .05$, $n_p^2 = .020$, no main effect of Belief, $F(1,46) = 2.17$, $p > .05$, $n_p^2 = .045$, and no interaction between Session and Belief, $F(1,46) = .241$, $p > .05$, $n_p^2 = .005$

(Figure 5c). Similarly, for Optimism Bias there was no main effect of Session, $F(1,46) = .398$, $p > .05$, $n_p^2 = .009$, no main effect of Belief, $F(1,46) = 1.09$, $p > .05$, $n_p^2 = .023$, and no interaction between Session and Belief, $F(1,46) = .030$, $p > .05$, $n_p^2 = .001$ (Figure 5d).

7. Exploratory correlations

Based on the Watching Eyes model (Conty et al., 2016), we proposed that audience effects may be mediated by an increase in self-referential processing when being seen. In order to test the relationship between these processes, we computed exploratory Pearson correlations between the measures obtained in the different tasks (self-referential processing, prosocial behaviour, confidence bias and optimism bias), and between questionnaire scores and task measures. None of the exploratory correlations was significant ($p > .05$ for all).

8. Discussion

The cognitive mechanisms by which being watched triggers changes in behaviour to signal good reputation (audience effects) are poorly understood. Here we proposed that these changes happen because the belief in being watched increases self-referential processing. Our study aimed to test this model by using a novel deceptive video-conference paradigm (Cañigueral & Hamilton, 2019), where participants either believed there was a real video-feed with a confederate or knew they were watching pre-recorded video-clips of another confederate. Results showed that, although there was a self-referential memory effect, it did not increase when participants believed they were being watched. We also failed to replicate previous findings showing that the belief in being watched increases prosocial behaviour, and similarly there was no effect of this manipulation on measures of self-awareness. Nonetheless, we have strong evidence that the deceptive video-conference manipulation was effective: participants in the ON setting rated the confederate as more likeable, natural and reciprocal than participants in the OFF setting. Based on previous evidence and these

findings, we identify key research areas that will help elucidating the mechanisms underlying audience effects.

8.1. Being watched and self-referential memory

To assess how self-referential memory is affected by the belief in being watched, participants completed a self-referential effect memory task, which measures their memory sensitivity to recall adjectives related to the self and to another person (Lombardo et al., 2007). Results showed that items related to the self were better recognised than items related to another person, across baseline and test session, for both ON and OFF group. This result proves that the task worked well when embedded in the deceptive video-conference setting. Contrary to our hypothesis, we did not find evidence that the belief in being watched increased self-referential processing. However, there was strong evidence that self- and other-related adjectives were better remembered in the baseline session than in the test session, both for ON and OFF group. This suggests that instead of a self-referential effect of someone watching us, the presence of a face (regardless of whether it could or could not see us) acted as a distractor: participants paid less attention to the adjectives and this impacted both its encoding and later recognition. Indeed, eye-tracking studies have shown that overt visual attention prioritises social information (e.g. faces) over non-social information, and that this happens reflexively (Rösler, End, & Gamer, 2017).

Our results do not corroborate those by Hietanen & Hietanen (2017), where they show that live direct gaze increases self-referential processing. A key difference between both studies is that in Hietanen & Hietanen (2017) participants were face-to-face with the confederate and experienced true direct gaze, whereas in our study participants interacted with the confederate through a screen that resembled a video-conference software. Although we designed our stimuli to match the ambiguous gaze pattern characteristic of video-conferences (where gaze is slightly averted when the other person is watching me), this

means that there was no true direct gaze. Thus, this could indicate that the belief in being watched *per se* is not enough to trigger self-referential processing, but rather needs to be embedded in true direct gaze. Another possible explanation is that the tasks used in both studies engage different cognitive processes. While completion of our task requires deep encoding of items for later recognition (Craik & Tulving, 1975), the pronoun-selection task used by Hietanen & Hietanen (2017) is more intuitive and has previously been shown to be sensitive to manipulations of self-awareness (Davis & Brock, 1975).

8.2. Being watched and prosocial behaviour

To assess how prosocial behaviour changes when being watched we used the Story task, which was found to be sensitive to the deceptive video-conference manipulation in a previous study using the same paradigm (Cañigueral & Hamilton, 2019). Unfortunately, these results are not replicated: prosociality of the answers does not change from baseline session to ON test session, and there is no difference between ON and OFF test sessions. Similarly, we could not replicate the correlation between social anxiety traits and change in prosocial behaviour from baseline to ON setting.

This lack of effect could be accounted for by differences in the cover story used in both studies. While in the previous study participants believed the confederate was a student volunteering in a charity (i.e. she was a positive example of prosocial behaviour), here they believed she was a PhD student working in the university, who had no explicit links to charity or volunteering work. It could be that the social context and the identity of the confederate is relevant for audience effects: participants might perceive someone linked to charitable work as more entitled to judge their actions than a random student, and the motivation to show that 'I'm prosocial' will be stronger for the former. For instance, low-status participants tend to be more prosocial than high-status participants (Guinote, Cotzia, Sandhu, & Siwa, 2015; Piff, Kraus, Côté, Cheng, & Keltner, 2010), and it has been suggested

that they do so to increase their social status in a high-status group (Kafashan, Sparks, Griskevicius, & Barclay, 2014) and, in turn, their reputation in the group. This suggests that the identity or social context of the observer in relation to the participant (e.g. social status) may be a strong modulator of audience effects on prosocial behaviour.

8.3. Being watched and self-awareness

Participants completed two tasks that measured self-awareness implicitly: the Confidence Bias task to measure confidence bias (metacognitive self-awareness; Fleming & Dolan, 2012; Harvey, 1997), and the Optimism bias questionnaire to measure the optimism bias (self-assessment; Sharot, 2011). Results showed there was no effect of the belief in being watched in either self-awareness task. These results are similar to those obtained in the self-referential effect memory task, but here performance from baseline session to test session did not decrease. This suggests that even when the task did not require deep encoding of information, and performance of participants was not negatively impacted by the social presence, self-awareness did not increase when being watched. A main limitation in these two tasks is that there was no video-feed of the confederate. Instead, participants were shown a still frame of the video-clip plus the message 'Screen Share active', indicating that the confederate could still see their answers. However, participants might have felt that it was ambiguous whether the confederate could only see their answers or could also see them, and this might have weakened the effect of the belief in being watched.

Another caveat is that different forms of self-awareness might have different sensitivity to the belief in being watched. It has been shown that direct gaze and the belief in being watched increase self-awareness of physiological signals in response to emotional pictures (Baltazar et al., 2014; Hazem et al., 2017). Instead, the tasks we use tap into metacognitive self-awareness and self-assessment, which require participants to reflect on their own judgements and self-knowledge. It could be that, compared to effects on bodily

self-awareness, effects on metacognitive self-awareness need stronger (or less ambiguous) belief manipulations. Thus, an interesting question is whether and how different forms of self-awareness are distinctly modulated by the belief in being watched embedded in eye gaze.

8.4. Implications and further research

These findings have important implications for future research on the cognitive mechanisms underlying audience effects. We show that our deceptive video-conference paradigm, which combines high ecological validity and experimental control, is successful in manipulating beliefs of participants (see also Cañigueral & Hamilton, 2019). This is supported by strong evidence showing that participants in the ON setting rated the confederate as more likeable, natural and reciprocal than participants in the OFF setting. However, our results indicate that the relationship between the belief in being watched, self-referential processing and subsequent behavioural effects (on prosocial behaviour and self-awareness) might not be as straightforward as we proposed. For instance, comparison with previous findings (Hietanen & Hietanen, 2017) suggests that self-referential processing might be differently modulated by subtle manipulations of true direct gaze. It also suggests that the belief in being watched might have different effects on distinct forms of self-referential processing (e.g. deep encoding of self-related items as used in the present study (Craik & Tulving, 1975) vs. intuitive pronoun-selection task used by Hietanen & Hietanen (2017)). Similarly, different forms of self-awareness may have different sensitivity to the belief in being watched: it could be that bodily self-awareness (Baltazar et al., 2014; Hazem et al., 2017) is more sensitive to audience manipulations than metacognitive self-awareness. Future studies that contrast audience effects on different forms of self-referential processing and self-awareness are critical to elucidate the role of the self in audience effects.

Moreover, the social context and the identity of the confederate may also be relevant for audience effects. Using the same Story task and deceptive video-conference paradigm

across two studies, we find that participants act more prosocially in the ON setting (compared to the OFF setting) if they believe the confederate is volunteering in a charity (Cañigueral & Hamilton, 2019) but not if she is presented as another student in the university (present study). In line with this, previous studies have shown that low-status individuals tend to be more prosocial, likely because this will help them increase their reputation in the group (Guinote et al., 2015; Kafashan et al., 2014; Piff et al., 2010). This suggests that participants not only process whether they are being seen or not, but also the identity of the observer in relation to them, and whether s/he poses a challenge to their reputation. Future studies could take a closer look at this question by systematically modulating the belief in being watched and social context of the study (e.g. identity of the observer).

Finally, it has been suggested that individual differences in public self-awareness and social anxiety modulate changes in prosocial behaviour when being watched (Cañigueral & Hamilton, 2019; Pfattheicher & Keller, 2015). Likewise, personality traits such as high prevention-focused self-regulation (i.e. tendency to ensure safety and security instead of striving for ideal gains and goals) increase prosocial cooperation when being watched (Keller & Pfattheicher, 2011). Although exploratory correlations between questionnaires scores (e.g. social anxiety traits, self-awareness) and task measures did not yield any significant relationship in the present study, future studies could directly test the role of personality traits in audience effects.

9. Conclusion

This study aimed to test whether audience effects related to reputation management (e.g. increase in prosocial behaviour when being watched) are mediated by an increase in self-referential processing. To do so, we used a novel deceptive video-conference paradigm (Cañigueral & Hamilton, 2019), where participants believe that video-clips of a confederate are a real video-feed or pre-recorded video-clips. Results show that both the deceptive belief

manipulation and the self-referential processing task were effective, but there was no influence of the belief in being watched on the latter. Equally, there was no effect of such manipulation on other measures of self-awareness and prosocial behaviour. Our findings indicate that the relationship between the belief in being watched, self-referential processing and subsequent behavioural effects (on prosocial behaviour and self-awareness) is not as straightforward as we hypothesised. We propose that further research on the role of the self, social context and personality traits will help elucidating the mechanisms underlying audience effects.

Declarations of interest. None.

Compliance with Ethical Standards. All procedures were approved by the UCL Research Ethics Committee and were in accordance with the Declaration of Helsinki and APA ethical standards.

Funding. This work was supported by the European Research Council (Starting Grant 313398-INTERACT). Roser Cañigueral acknowledges financial support from “la Caixa” Foundation (ID 100010434, grant code LCF/BQ/EU16/11560039). The funding bodies had no involvement in the execution of this study and writing of the report.

References

Anderson, N. H. (1968). Likableness Ratings of 555 Personality-Trait Words. *Journal of Personality and Social Psychology*, *9*(3), 272–279.

Bagby, A. R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*, 23–32.

Baltazar, M., Hazem, N., Vilarem, E., Beaucousin, V., Picq, J. L., & Conty, L. (2014). Eye contact elicits bodily self-awareness in human adults. *Cognition*, *133*(1), 120–127.

Baron-Cohen, S., Wheelwright, S. J., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414.

Bereczkei, T., Birkas, B., & Kerekes, Z. (2007). Public charity offer as a proximate factor of evolved reputation-building strategy: an experimental analysis of a real-life situation. *Evolution and Human Behavior*, *28*(4), 277–284. https://doi.org/10.1016/j.evolhumbehav.2007.04.002

Bond, C. F. J. (1982). Social Facilitation: A Self-Presentational View. *Journal of Personality and Social Psychology*, *42*(6), 1042–1050.

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, *285*(20180116). https://doi.org/10.1098/rspb.2018.0116

Cage, E. A. (2015). *Mechanisms of social influence: Reputation management in typical and autistic individuals.*

Cage, E. A., Pellicano, E., Shah, P., & Bird, G. (2013). Reputation management: Evidence for ability but reduced propensity in autism. *Autism Research*, *6*(5), 433–442.

Cañigueral, R., & Hamilton, A. F. de C. (2017). Effects of being watched on self-referential processing, self-awareness and prosocial behaviour. Retrieved July 31, 2017, from osf.io/xtmh8

Cañigueral, R., & Hamilton, A. F. de C. (2019). Being watched: Effects of an audience on eye gaze and prosocial behaviour. *Acta Psychologica*, *195*, 50–63. https://doi.org/10.1016/j.actpsy.2019.02.002

Conty, L., George, N., & Hietanen, J. K. (2016). Watching Eyes effects: when others meet the self. *Consciousness and Cognition*, *45*, 184–197.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294.

Davis, D., & Brock, T. C. (1975). Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology*, *11*(4), 381–388.

Emler, N. (1990). A social psychology of reputation. *European Review of Social Psychology*, *1*(1), 171–193.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, *43*(4), 522–527.

Filiz-Ozbay, E., & Ozbay, E. Y. (2014). Effect of an audience in public goods provision. *Experimental Economics*, *17*(2), 200–214. https://doi.org/10.1007/s10683-013-9363-y

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1338–1349.

Fridlund, A. J. (1991). Sociality of Solitary Smiling: Potentiation by an Implicit Audience.

Journal of Personality and Social Psychology, *60*(2), 229–240.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Fullwood, C., & Doherty-Sneddon, G. (2006). Effect of gazing at the camera during a video link on recall. *Applied Ergonomics*, *37*(2), 167–175. https://doi.org/10.1016/j.apergo.2005.05.003

Guala, F., & Mittone, L. (2010). Paradigmatic experiments: The Dictator Game. *Journal of Socio-Economics*, *39*(5), 578–584. https://doi.org/10.1016/j.socec.2009.05.007

Guinote, A., Cotzia, I., Sandhu, S., & Siwa, P. (2015). Social status modulates prosocial behavior and egalitarianism in preschool children and adults. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(3), 731–736. https://doi.org/10.1073/pnas.1414550112

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity an anonymous economic game. *Evolution and Human Behavior*, *26*(3), 245–256.

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Neuroscience*, *1*(2), 78–82.

Hazem, N., George, N., Baltazar, M., & Conty, L. (2017). I know you can see me: Social attention influences bodily self-awareness. *Biological Psychology*, *124*, 21–29.

Hietanen, J. O., & Hietanen, J. K. (2017). Genuine eye contact elicits self-referential processing. *Consciousness and Cognition*, *51*, 100–115.

Hull, J. G., Van Treuren, R. R., Ashford, S. J., Propsom, P., & Andrus, B. W. (1988). Self-Consciousness and the Processing of Self-Relevant Information. *Journal of Personality and Social Psychology*, *54*(3), 452–465.

Izuma, K., Matsumoto, K., Camerer, C. F., & Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences*, *108*(42), 17302–17307.

Izuma, K., Saito, D. N., & Sadato, N. (2009). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of Cognitive Neuroscience*, *22*(4), 621–631.

Izuma, K., Saito, D. N., & Sadato, N. (2010). The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience*, *5*(2), 133–147.

Kafashan, S., Sparks, A., Griskevicius, V., & Barclay, P. (2014). Prosocial Behavior and Social Status. In J. T. Cheng, J. L. Tracy, & C. Anderson (Eds.), *The Psychology of Social Status* (pp. 139–158). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-0867-7

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the Assumptions of Economics. *The Journal of Business*, *59*(4).

Keller, J., & Pfattheicher, S. (2011). Vigilant Self-Regulation, Cues of Being Watched and Cooperativeness. *European Journal of Personality*, *25*, 363–372. https://doi.org/10.1002/per.797

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–435. https://doi.org/10.1006/jhev.2001.0468

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, *340*, 294–340.

Liebowitz, M. R. (1987). Liebowitz Social Anxiety Scale. *Modern Problems of Pharmapsychiatry*, *22*, 141–173.

Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in austism. *PLoS ONE*, *2*(9).

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., … Baron-Cohen, S. (2010). Atypical neural self-representation in autism. *Brain*,

133(2), 611–624. https://doi.org/10.1093/brain/awp306

Mansour, H., & Kuhn, G. (2019). Studying "natural" eye movements in an "unnatural" social environment: The influence of social activity, framing, and sub-clinical traits on gaze aversion. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021818819094

Pfattheicher, S., & Keller, J. (2015). The watching eyes phenomenon: The role of a sense of being seen and public self-awareness. *European Journal of Personality*, *45*(5), 560–566. https://doi.org/10.1002/ejsp.2122

Pfeiffer, T., & Nowak, M. A. (2006). All in the game. *Nature*, *441*(June), 583–584. https://doi.org/10.1038/441583a

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having Less, Giving More: The Influence of Social Class on Prosocial Behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784. https://doi.org/10.1037/a0020092

Pönkänen, L. M., Peltola, M. J., & Hietanen, J. K. (2011). The observer observed: Frontal EEG asymmetry and autonomic responses differentiate between another person's direct and averted gaze when the face is seen live. *International Journal of Psychophysiology*, *82*(2), 180–187. https://doi.org/10.1016/j.ijpsycho.2011.08.006

Raihani, N. J., & Smith, S. (2015). Competitive Helping in Online Giving. *Current Biology*, *25*(9), 1183–1186. https://doi.org/10.1016/j.cub.2015.02.042

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, *9*(2), 79–101.

Rösler, L., End, A., & Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLoS ONE*, *12*(7), 1–14.

Satow, K. L. (1975). Social Approval and Helping. *Journal of Experimental Social Psychology*, *11*(6), 501–509. https://doi.org/10.1016/0022-1031(75)90001-3

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition*, *12*(1), 127–144.

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), 941–945. https://doi.org/10.1016/j.cub.2011.10.030

Silver, I. M., & Shaw, A. (2018). Pint-Sized Public Relations: The Development of Reputation Management. *Trends in Cognitive Sciences*, *22*(4), 277–279. https://doi.org/10.1016/j.tics.2018.01.006

Symons, C. S., & Johnson, B. T. (1997). The Self-Reference Effect in Memory: A Meta-Analysis. *Psychological Bulletin*, *121*(3), 371–394.

Tennie, C., Frith, U., & Frith, C. D. (2010). Reputation management in the age of the world-wide web. *Trends in Cognitive Sciences*, *14*(11), 482–488.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*(8), 376–382. https://doi.org/10.1016/j.tics.2010.05.005

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, *9*(4), 507–533.

von Grünau, M., & Anston, C. (1995). The detection of gaze direction: A stare-in-the-crowd effect. *Perception*, *24*(11), 1297–1313. https://doi.org/10.1068/p241297

Zajonc, R. B. (1965). Social Facilitation. *Science*, *149*(3681), 269–274.

Figures

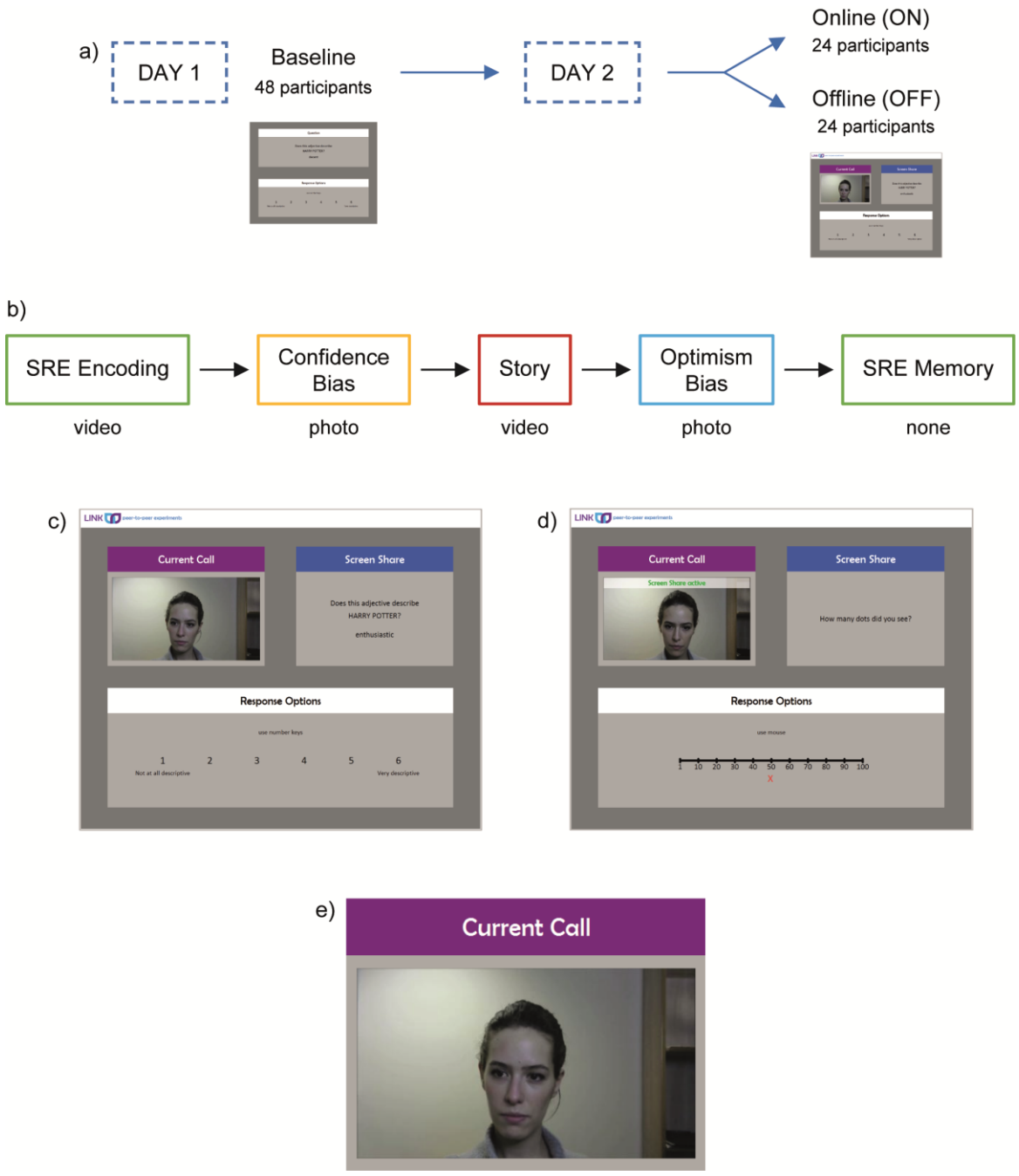

Figure 1. a) Overview of the study over the two days. b) Procedure of the study and type of stimuli used in each task. c-d) Screenshots of LINK during a task with a video-clip (c) and a picture (d) in the ON condition. For the OFF condition the top boxes were called Videos and Side Screen, respectively. e) Zoomed screenshot of the confederate with slightly averted gaze to make participants believe she is watching them and their choices.

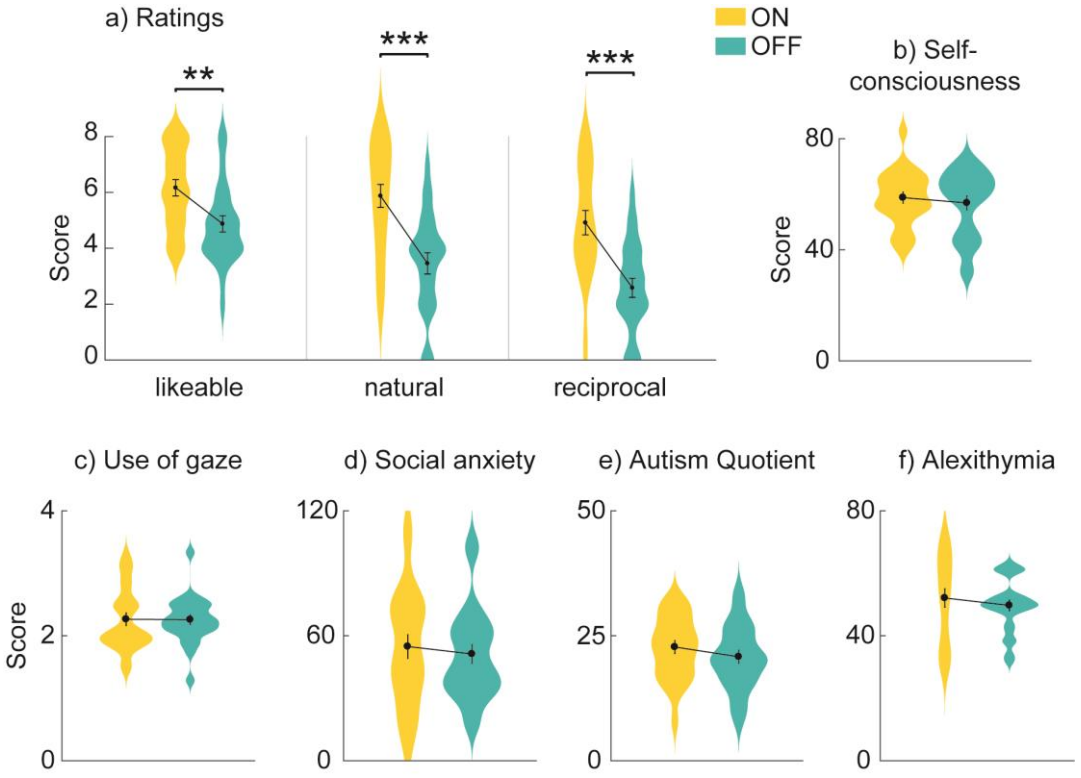

Figure 2. Mean (filled circle), SE (error bars), and frequency of values (width of distribution) of scores in the questionnaires. Asterisks signify difference between ON and OFF setting at $p < .1$ (+), $p < .05$ (*), $p < .01$ (**) and $p < .001$ (***).

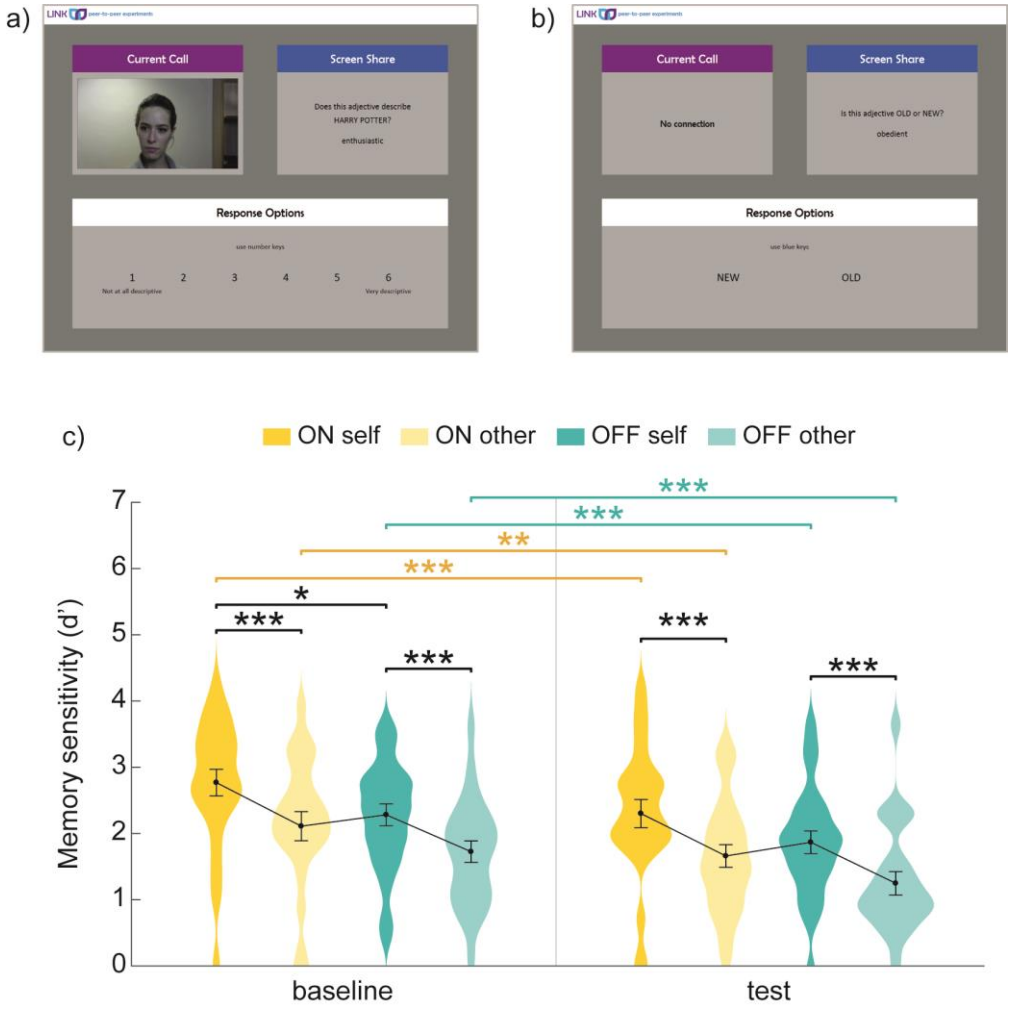

Figure 3. a) Encoding phase of SRE task during ON condition. Screen Share shows question 'Does this adjective describe HARRY POTTER? - enthusiastic'. b) Memory phase of SRE task during ON condition. Screen Share shows question 'Is this adjective OLD or NEW? - obedient'. c) Mean (filled circle), SE (error bars), and frequency of values (width of distribution) for memory sensitivity. Asterisks signify difference between ON and OFF setting at $p < .1$ (+), $p < .05$ (*), $p < .01$ (**) and $p < .001$ (***).

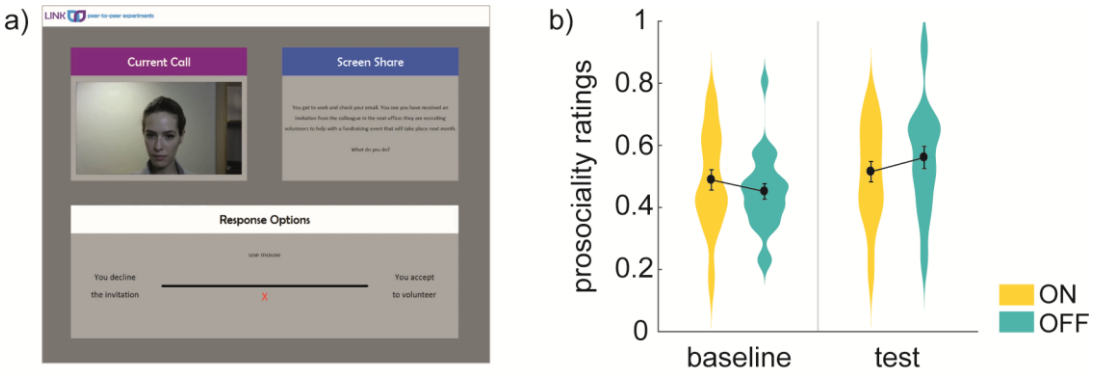

Figure 4. a) Story task during ON condition. Screen Share shows the dilemma, and the Reponse Options box shows the two possible answers. b) Mean (filled circle), SE (error bars), and frequency of values (width of distribution) for prosociality ratings.

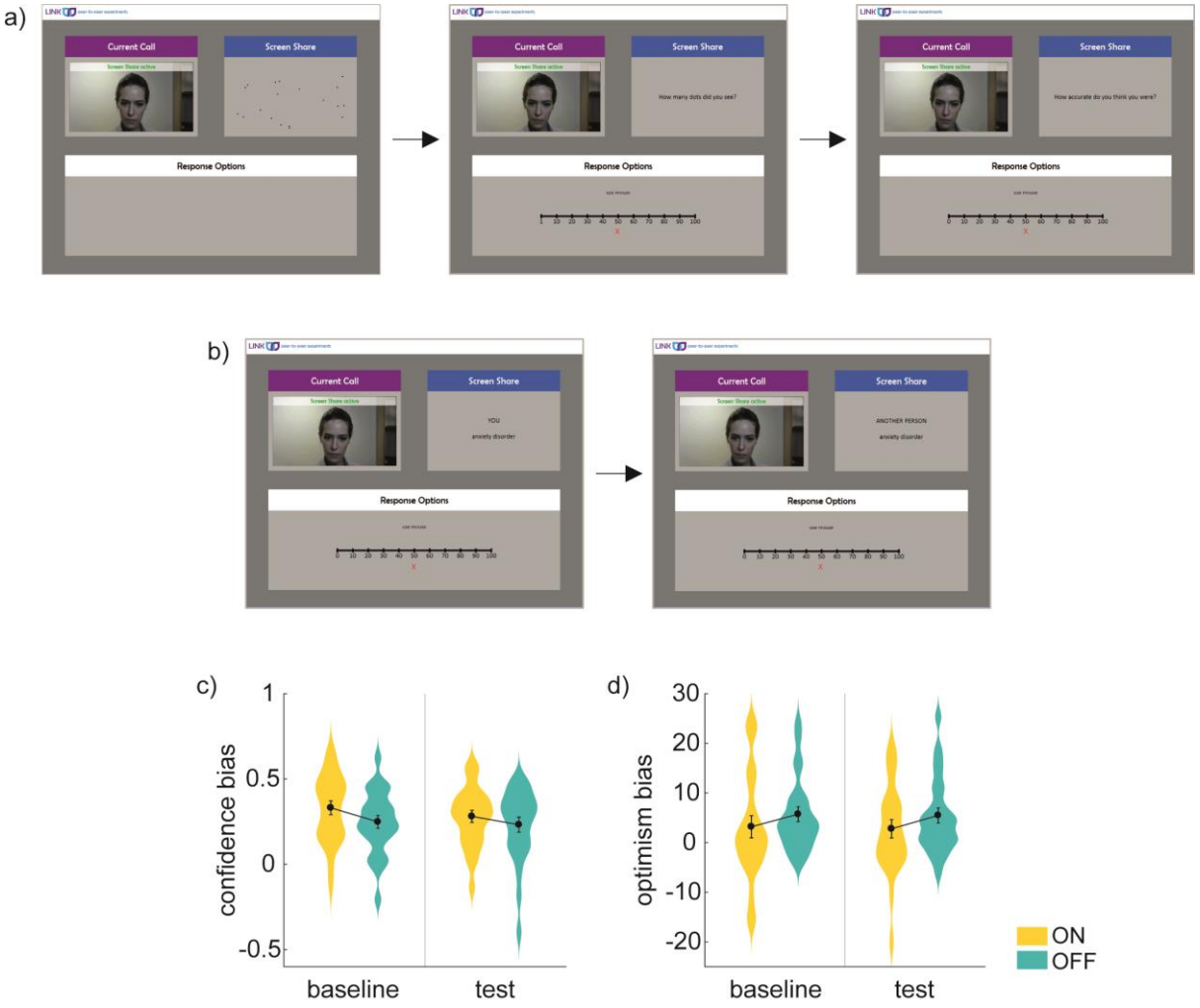

Figure 5. a) Confidence Bias task during ON condition. Screen Share first shows the dots, followed by the questions 'How many dots did you see?' and 'How accurate you think you were?'. b) Optimism Bias questionnaire during ON condition. Screen Share first shows 'YOU - anxiety disorder', followed by 'ANOTHER PERSON - anxiety disorder'. c-d) Mean (filled circle), SE (error bars), and frequency of values (width of distribution) for confidence (c) and optimism (d) bias.

Supplementary materials

S1. Self-consciousness scale

Private self-consciousness

I'm always trying to figure myself out. (1)

Generally, I'm not very aware of myself. (3)[b]

I reflect about myself a lot. (5)

I'm often the subject of my own fantasies. (7)

I never scrutinize myself. (9)[b]

I’m generally attentive to my inner feelings. (13)

I'm constantly examining my motives. (15)

I sometimes have the feeling that I'm off somewhere watching myself. (18)

I'm alert to changes in my mood. (20)

I’m aware of the way my mind works when I work through a problem. (22)

Public self-consciousness

I'm concerned about my style of doing things. (2)

I'm concerned about the way I present myself. (6)

I'm self-conscious about the way I look. (11)

l usually worry about making a good impression. (14)

One of the last things I do before I leave my house is look in the mirror. (17)

I’m concerned about what other people think of me. (19)

I’m usually aware of my appearance. (21)

Social Anxiety

It takes me time to overcome my shyness in new situations. (4)

I have trouble working when someone is watching me. (8)

I get embarrassed very easily. (10)

I don't find it hard to talk to strangers. (12)[b]

I feel anxious when I speak in front of a group. (16)

Large groups make me nervous. (23)

(#) = sequence of items in questionnaire

b = item reversed for scoring

S2. Gaze questionnaire

1. It is easy for me to decide how much eye contact is appropriate.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

2. If I want to know how someone feels, then I look at their eyes

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

3. I notice when people are looking at me.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

4. I am not sure how long I should look at someone's eyes when talking to them.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

5. I like to be the centre of attention.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

6. I understand someone's emotions more if they look at me.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

7. When I am speaking to someone, I deliberately move my eyes in a particular pattern or look at a particular place.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

8. I feel anxious if someone looks directly at my eyes.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

9. I need to think about whether or not to make eye-contact.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

10. I like to stare at someone until that person looks away.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

11. If I want to know what someone's intentions are, then I look at their eyes.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

12. I feel uncertain or confused if someone looks directly at my eyes.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

13. As a child or young person, I was told to look at people's eyes more often during conversations.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

14. I prefer to sit next to someone rather than opposite them to avoid eye contact.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

15. Sometimes I feel like everyone is staring at me.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

16. I do not deliberately control where I am looking during a conversation.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

17. I understand someone's thoughts more if they look at me

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

18. If I want to know what someone is thinking, then I look at their eyes

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

19. I find eye-contact intense and overwhelming, like looking straight at a very bright light.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

20. As a child I was never taught about eye-contact.

strongly disagree | slightly disagree | neither agree nor disagree | slightly agree | strongly agree

S3. Liebowitz Social Anxiety Scale

Fear/Anxiety:		Avoidance:	
0 = None	2 = Moderate	0 = Never (0%)	2 = Often (33-67%)
1 = Mild	3 = Severe	1 = Occasionally (1-33%)	3 = Usually (67-100%)

	Fear/Anxiety	Avoidance
1. Telephoning in public.		
2. Participating in small groups.		
3. Eating in public places.		
4. Drinking with others in public places.		
5. Talking to people in authority.		
6. Acting, performing or giving a talk in front of an audience.		
7. Going to a party.		
8. Working while being observed.		
9. Writing while being observed.		
10. Calling someone you don't know very well.		
11. Talking with people you don't know very well.		
12. Meeting strangers.		
13. Urinating in a public bathroom.		
14. Entering a room when others are already seated.		
15. Being the center of attention.		
16. Speaking up at a meeting.		
17. Taking a test.		
18. Expressing disagreement/disapproval to people you don't know very well.		
19. Looking at people you don't know very well in the eyes.		
20. Giving a report to a group.		
21. Trying to pick up someone.		
22. Returning goods to a store.		
23. Giving a party.		
24. Resisting a high pressure salesperson.		

S4. Autism Quotient

1. I prefer to do things with others rather than on my own.	definitely agree	slightly agree	slightly disagree	definitely disagree
2. I prefer to do things the same way over and over again.	definitely agree	slightly agree	slightly disagree	definitely disagree
3. If I try to imagine something, I find it very easy to create a picture in my mind.	definitely agree	slightly agree	slightly disagree	definitely disagree
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.	definitely agree	slightly agree	slightly disagree	definitely disagree

5. I often notice small sounds when others do not.	definitely agree	slightly agree	slightly disagree	definitely disagree
6. I usually notice car number plates or similar strings of information.	definitely agree	slightly agree	slightly disagree	definitely disagree
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.	definitely agree	slightly agree	slightly disagree	definitely disagree
8. When I'm reading a story, I can easily imagine what the characters might look like.	definitely agree	slightly agree	slightly disagree	definitely disagree
9. I am fascinated by dates.	definitely agree	slightly agree	slightly disagree	definitely disagree
10. In a social group, I can easily keep track of several different people's conversations.	definitely agree	slightly agree	slightly disagree	definitely disagree
11. I find social situations easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
12. I tend to notice details that others do not.	definitely agree	slightly agree	slightly disagree	definitely disagree
13. I would rather go to a library than a party.	definitely agree	slightly agree	slightly disagree	definitely disagree
14. I find making up stories easy.	definitely agree	slightly agree	slightly disagree	definitely disagree
15. I find myself drawn more strongly to people than to things.	definitely agree	slightly agree	slightly disagree	definitely disagree
16. I tend to have very strong interests which I get upset about if I can't pursue.	definitely agree	slightly agree	slightly disagree	definitely disagree
17. I enjoy social chit-chat.	definitely agree	slightly agree	slightly disagree	definitely disagree
18. When I talk, it isn't always easy for others to get a word in edgeways.	definitely agree	slightly agree	slightly disagree	definitely disagree
19. I am fascinated by numbers.	definitely agree	slightly agree	slightly disagree	definitely disagree
20. When I'm reading a story, I find it difficult to work out the characters' intentions.	definitely agree	slightly agree	slightly disagree	definitely disagree
21. I don't particularly enjoy reading fiction.	definitely agree	slightly agree	slightly disagree	definitely disagree
22. I find it hard to make new friends.	definitely agree	slightly agree	slightly disagree	definitely disagree
23. I notice patterns in things all the time.	definitely agree	slightly agree	slightly disagree	definitely disagree

24. I would rather go to the theatre than a museum.	definitely agree	slightly agree	slightly disagree	definitely disagree
25. It does not upset me if my daily routine is disturbed.	definitely agree	slightly agree	slightly disagree	definitely disagree
26. I frequently find that I don't know how to keep a conversation going.	definitely agree	slightly agree	slightly disagree	definitely disagree
27. I find it easy to "read between the lines" when someone is talking to me.	definitely agree	slightly agree	slightly disagree	definitely disagree
28. I usually concentrate more on the whole picture, rather than the small details.	definitely agree	slightly agree	slightly disagree	definitely disagree
29. I am not very good at remembering phone numbers.	definitely agree	slightly agree	slightly disagree	definitely disagree
30. I don't usually notice small changes in a situation, or a person's appearance.	definitely agree	slightly agree	slightly disagree	definitely disagree
31. I know how to tell if someone listening to me is getting bored.	definitely agree	slightly agree	slightly disagree	definitely disagree
32. I find it easy to do more than one thing at once.	definitely agree	slightly agree	slightly disagree	definitely disagree
33. When I talk on the phone, I'm not sure when it's my turn to speak.	definitely agree	slightly agree	slightly disagree	definitely disagree
34. I enjoy doing things spontaneously.	definitely agree	slightly agree	slightly disagree	definitely disagree
35. I am often the last to understand the point of a joke.	definitely agree	slightly agree	slightly disagree	definitely disagree
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.	definitely agree	slightly agree	slightly disagree	definitely disagree
37. If there is an interruption, I can switch back to what I was doing very quickly.	definitely agree	slightly agree	slightly disagree	definitely disagree
38. I am good at social chit-chat.	definitely agree	slightly agree	slightly disagree	definitely disagree
39. People often tell me that I keep going on and on about the same thing.	definitely agree	slightly agree	slightly disagree	definitely disagree
40. When I was young, I used to enjoy playing games involving pretending with other children.	definitely agree	slightly agree	slightly disagree	definitely disagree
41. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.).	definitely agree	slightly agree	slightly disagree	definitely disagree
42. I find it difficult to imagine what it would be like to be someone else.	definitely agree	slightly agree	slightly disagree	definitely disagree

43. I like to plan any activities I participate in carefully.	definitely agree	slightly agree	slightly disagree	definitely disagree
44. I enjoy social occasions.	definitely agree	slightly agree	slightly disagree	definitely disagree
45. I find it difficult to work out people's intentions.	definitely agree	slightly agree	slightly disagree	definitely disagree
46. New situations make me anxious.	definitely agree	slightly agree	slightly disagree	definitely disagree
47. I enjoy meeting new people.	definitely agree	slightly agree	slightly disagree	definitely disagree
48. I am a good diplomat.	definitely agree	slightly agree	slightly disagree	definitely disagree
49. I am not very good at remembering people's date of birth.	definitely agree	slightly agree	slightly disagree	definitely disagree
50. I find it very easy to play games with children that involve pretending.	definitely agree	slightly agree	slightly disagree	definitely disagree

S5. Toronto Alexithymia Scale

Indicate how much you agree or disagree with each of the following statements. Just tick the appropriate box. **Use the middle box** ('I neither agree or disagree') **only if you are really unable to assess your behaviour.**	I strongly disagree	I quite disagree	I neither agree nor disagree	I quite agree	I strongly agree
1- I am often confused about what emotion I am feeling					
2- It is difficult for me to find the right words for my feelings					
3- I have physical sensations that even doctors don't understand					
4- I am able to describe my feelings easily					
5- I prefer to analyze problems rather than just describe them					
6- When I am upset, I don't know if I am sad, frightened, or angry					
7- I am often puzzled by sensations in my body					
8- I prefer to just let things happen rather than to understand why they turned out that way					
9- I have feelings that I can't quite identify					

10- Being in touch with emotions is essential					
11- I find it hard to describe how I feel about people					
12- People tell me to describe my feelings more					
13- I don't know what's going on inside me					
14- I often don't know why I am angry					
15- I prefer talking to people about their daily activities rather then their feelings					
16- I prefer to watch « light » entertainment shows rather than psychological dramas					
17- It is difficult for me to reveal my innermost feelings, even to close friends					
18- I can feel close to someone, even in moments of silence					
19- I find examination of my feelings useful in solving personal problems					
20- Looking for hidden meanings in movies or plays distracts from their enjoyment					

S6. Conversation with Alice

Experimenter presses "enter" to connect with student, and video of Alice appears.

Experimenter (E): Hi Alice, how're you? Can you hear me?

Alice (A): Hi! Yes I hear you; there's a bit of noise, but it's fine.

E: Yeah? Great, and can you see our participant here today?

A: Yes, hi!

E: Ok, so Alice, this is [name of participant]. [Name of participant] this is Alice…

A *(waving her hand)*: Hi, nice to meet you!

E: Now we need to check that the Screen Share is working… *(press number 5)* Can you tell me what number is on the Screen Share now, if you can see it?

A: Yes, number 5.

E: And now? *(press number 3)*

A: Hmm, 3.

E: Cool, it seems that everything's working well… So for the first half of the study [name of participant] will complete the Adjectives task [SRE Encoding phase task] and the Counting task [Confidence Bias task]. Alice, as you know for the Counting task we won't be able to have the video-feed, but you can still track the answers on the Screen Share. And whenever there's the video-feed active, please remember not to make any facial expression or say anything that could influence the participant's choices. Is everything clear? *(Inbetween Alice nods and smiles)*

A: Yes, everything's clear.

E: Great, are you ready then to start?

A: Yes, I'm ready!

Participant completes SRE Encoding phase task and Confidence Bias task, calls the experimenter, and she presses "space".

A: Hey, I'm ready for the next task!

E: Great, so the next task is the Story task. Alice, you will read the statement on the Screen Share to [name of participant] and ask him/her "what do you do?". Please, remember to keep your face neutral. Then, the last task will be the Predictions task [Optimism Bias questionnaire], and again you will only share the Screen Share for this one. *(Inbetween Alice nods and smiles)*

A: Yes, OK.

Participant completes Story task and Optimism Bias questionnaire, calls the experimenter, and she presses "space".

A: Well, thank you for doing the task! Speak to you later, [name of experimenter]. Bye!

E: Thank you Alice, speak to you, bye!

S7. Post-test questionnaire

Section 1

I liked Alice very much.

(disagree) 0 1 2 3 4 5 6 7 8 (agree)

I think the interaction with Alice was very natural.

(disagree) 0 1 2 3 4 5 6 7 8 (agree)

I think the interaction with Alice was very reciprocal.

(disagree) 0 1 2 3 4 5 6 7 8 (agree)

I think it is very important to donate money to charity.

(disagree) 0 1 2 3 4 5 6 7 8 (agree)

I think it is very important to do some voluntary work.

(disagree) 0 1 2 3 4 5 6 7 8 (agree)

Section 2

What do you think was the purpose of the experiment?

_____

_____

Please, explain if you followed any strategy when giving an answer on the…

Adjectives task? _____

_____

_____

Dots task? _____

_____

_____

Story task? _____

_____

_____

Predictions task? _____

_____

_____

Recognition task? _____

_____

_____

S8. Counterbalancing Conditions

Condition	Baseline session	Test session
1	Story 1	ON, Story 2, Confederate 1
2	Story 1	ON, Story 2, Confederate 2
3	Story 1	OFF, Story 2, Confederate 1
4	Story 1	OFF, Story 2, Confederate 2
5	Story 2	ON, Story 1, Confederate 1
6	Story 2	ON, Story 1, Confederate 2
7	Story 2	OFF, Story 1, Confederate 1
8	Story 2	OFF, Story 1, Confederate 2

ON=online setting; OFF= offline setting

S9. Adjectives

Baseline session

Self (30):

modest	polite	authoritative
cordial	efficient	old-fashioned
loyal	creative	unpleasant
relaxed	active	forgetful
self-critical	tolerant	clumsy
talkative	disagreeable	indecisive
open-minded	complaining	demanding
kind	suspicious	unhealthy
happy	gossipy	dominating
clever	strict	nervous

Other (30):

charming	skillful	clear-headed
decent	easygoing	clean
truthful	innocent	friendly

brilliant	prejudiced	insecure
helpful	depressed	passive
talented	deceptive	imitative
sensible	hesitant	submissive
gentle	discriminating	obstinate
amusing	extravagant	unpunctual
disobedient	impolite	messy

Distracter (60):

considerate	artistic	inconsistent
kind-hearted	precise	disturbed
responsible	social	inefficient
warm-hearted	comical	uninspiring
trustful	convincing	unsympathetic
honorable	meditative	hot-tempered
grateful	lucky	irritable
smart	perfectionistic	careless
respectful	well-spoken	boastful
original	outstanding	vain
constructive	radical	argumentative
sympathetic	anxious	bossy
productive	lonely	opportunist
neat	timid	shy
logical	immodest	unlucky
entertaining	tense	rebellious
romantic	worrying	daredevil
curious	sarcastic	inexperienced
positive	mediocre	preoccupied
skilled	stubborn	resigned

Test session

Self (30):

ingenious	experienced	frank
energetic	intelligent	optimistic

popular
competent
sincere
moral
thoughtful
wise
reliable
patient
intellectual
untidy
noisy
oversensitive
showy
frustrated
petty
pessimistic
weak
nonconfident
negligent
incompetent
reserved
impulsive
unappreciative
unfair

Other (30):

generous
enthusiastic
inventive
understanding
nice
adventurous
practical
proficient
honest
tender
mature
warm
interesting
prudent
cooperative
possessive
moody
overconfident
angry
cynical
lazy
unattentive
sad
antisocial
neurotic
superficial
prideful
aggressive
materialistic
childish

Distracter (60):

trustworthy
good-humored
educated
broad-minded
cheerful
reasonable
pleasant
bright
forgiving
admirable
attentive
realistic
progressive
good
accurate
agreeable
rational
modern
confident
calm
decisive
tidy
careful
disciplined
obedient
sentimental
fearless
sophisticated
unselfish
likable
choosy
troubled
tough
unskilled
ungraceful
silly
withdrawn
compulsive
unhappy

fearful	irrational	blunt
superstitious	foolish	self-concerned
pompous	helpless	eccentric
illogical	dull	skeptical
unproductive	hypochondriac	undecided
overcritical	aimless	unpopular
resentful	satirical	clownish

S10. Stories

Story 1

It’s Monday morning. You leave home and head toward the tube station to go to work. You are almost arriving to the platform when you hear the beeps announcing the tube's doors will close. What do you do? You run and catch the tube / You wait for the next one

You get to work and check your email. You see you have received an invitation from the colleague in the next office: they are recruiting volunteers to help with a fundraising event that will take place next month. What do you do? You decline the invitation / You accept to volunteer

At noon you go out to a nearby restaurant to have lunch. When you pay the waitress gives you the change, but there's more than should be. What do you do? You tell her the change is wrong / You don't say anything

After lunch you still have a lot of work to do, but you want to leave early this afternoon because you have planned to go to an art exhibition. However, you receive a call from a colleague: you need to discuss some issues related to a project, but she keeps chatting about an argument she had with her partner. What do you do? You keep trying to comfort her / You change the topic to discuss the project

In the end you have enough time to visit the art exhibition. Before leaving, you see a couple of collection boxes asking for a donation to help cover the costs of the exhibition. What do you do? You continue your way out / You donate something

On your way back home, you see a homeless man asking for money. He looks at you and asks if you can give him some coins. What do you do? You give him some money / You continue your way back home

Story 2

It's Friday afternoon and you're working hard to finish your essay before tomorrow, since a friend is arriving to visit you for the weekend. However, your friend John calls you to invite you to the cinema this evening: he had a date with a girl and had bought tickets, but she just cancelled it. What do you do? You go to the cinema / You tell him you are busy

The next morning you go to the train station to pick up your friend. While you wait for her, you check your Facebook on the cell phone and see a post from your flatmate's friend: he's asking for volunteers to help taking care of disabled children in the school where he works. What do you do? You continue checking posts / You say you'd like to help

It seems the train has been delayed, so you decide to have a walk outside the station. Right outside the station you see a homeless man juggling to music. When he finishes, he asks you for money. What do you do? You go back to the station / You give him some money

Finally, the train arrives and you meet your friend. You need to take a bus to go back home and leave the luggage, and you know there is one leaving from the far side of the station in 5 minutes. What do you do? You run to the bus stop / You wait for the next one

Then, you go to a pub to have a drink while you decide what to do. Your friend takes a seat and you go to the bar to order. When you pay, you realise the barman has given you more change than he should have done. What do you do? You tell him the change is wrong / You don't say anything

Finally, you decide to visit a museum. Although the entrance is free, there is a collection box to donate something to maintain the museum. What do you do? You donate something / You don't donate

S11. Items for predictions

fraud when buying something on the internet
card fraud
household accident
mouse/rat in house
more than £30000 debts
miss a flight
death before 80
witness a traumatising accident
domestic burglary
bone fracture
depression
heart failure
obesity
diabetes (type 2)
victim of violence by stranger
disease of spinal cord
serious hearing problems
infertility
dementia
drug abuse
being convicted of crime
house vandalised
gluten intolerance
appendicitis
age related blindness
death before 60
alcoholism
Parkinson's disease
back pain
being fired
eye cataract (clouding of the lens of the eye)
skin burn
hospital stay longer than three weeks
victim of bullying at work (nonphysical)
theft from person
sexual dysfunction
hepatitis A or B
severe teeth problems when old
cancer (colon/lung/prostate/breast/skin)
abnormal heart rhythm
victim of violence by acquaintance
herpes
migraine
having a stroke
victim of violence at home
severe insomnia
death before 70
severe injury due to accident (traffic or house)
autoimmune disease
victim of mugging
asthma
blood clot in vein
ulcer
kidney stones
Alzheimer's disease
anxiety disorder
limb amputation
epilepsy
liver disease
death by infectio

Supplementary tables

Table S1. Descriptives for post-test ratings and questionnaires

	Measure	ON	OFF
Ratings	likeable	*M* = 6.17 *SD* = 1.39	*M* = 4.88 *SD* = 1.42
	natural	*M* = 5.88 *SD* = 1.95	*M* = 3.46 *SD* = 1.72
	reciprocal	*M* = 4.92 *SD* = 2.13	*M* = 2.58 *SD* = 1.57
Questionnaires	self-consciousness	*M = 58.79* *SD* = 9.95	*M* = 56.88 *SD* = 12.23
	use of gaze	*M* = 2.27 *SD* = .499	*M* = 2.26 *SD* = .38
	social anxiety	*M* = 54.83 *SD = 27.86*	*M* = 51.25 *SD* = 22.05
	autism quotient	*M* = 22.75 *SD* = 6.33	*M* = 20.79 *SD* = 6.49
	alexithymia	*M* = 52.08 *SD* = 14.50	*M* = 49.71 *SD* = 8.48

Table S2. Descriptives for memory sensitivity (d')

Session	Target	ON	OFF
Baseline	Self	*M* = 2.77 *SD* = .985	*M* = 2.28 *SD* = .822
	Other	*M* = 2.11 *SD* = 1.10	*M* = 1.72 *SD* = .814
Test	Self	*M* = 2.30 *SD* = 1.06	*M* = 1.87 *SD* = .856
	Other	*M* = 1.66 *SD* = .854	*M* = 1.26 *SD = .886*

Table S3. Descriptives for task measures

Measure	Session	ON	OFF
Self bias	Baseline	*M* = .658 *SD* = .448	*M* = .557 *SD* = .536
	Test	*M* = .640 *SD* = .634	*M* = .619 *SD* = .401
Prosocial ratings	Baseline	*M* = .488 *SD* = .162	*M* = .451 *SD* = .116
	Test	*M* = .515 *SD* = .166	*M* = .561 *SD* = *.177*
Confidence bias	Baseline	*M* = .331 *SD* = .200	*M* = .249 *SD* = .191
	Test	*M* = .281 *SD* = .172	*M* = .232 *SD* = .219
Optimism bias	Baseline	*M* = 3.19 *SD* = 10.9	*M* = 5.72 *SD* = 7.53
	Test	*M* = 2.78 *SD* = 9.18	*M* = 5.48 *SD* = 7.69