



**HOW SOCIAL CONTEXTS AFFECT THE FORMATION OF FALSE  
MEMORIES AND THE MODERATING ROLE OF PROSOCIAL  
TENDENCIES**

ANNA MIA MANSTEIN

Orientador de Dissertação:  
DR. TERESA TEIXEIRA DE M. GARCIA MARQUES

Professor de Seminário de Dissertação:  
DR. TERESA TEIXEIRA DE M. GARCIA MARQUES

Dissertation submitted in partial fulfillment of the requirements for the degree of Master  
of Science in Social and Organizational Psychology

## Abstract

The following study is an investigation of context dependency in the formation of false memories and test if a prosocial context favors the formation of false memories compared to a selfish context. Furthermore, the prosociality of each participant was used as a moderator variable. In this study, false memories are formed by previously disseminated misinformation within the setup, based on the misinformation paradigm (Loftus et al., 1978). A total of 45 participants is presented with a pre-coded text about pro-social or selfish behavior, then watch one of two slide shows, followed by 50 sentences which contain misinformation, and complete a final memory test. A pro-social context is expected to lead to more false memories than a selfish context, supporting previous findings of consistency with human beliefs and supporting the assumption that humans are pro-social by nature. Results show no significant main effect of condition in the creation of false memories. However, interaction with the moderator reveals unexpected trends, suggesting that context influenced the formation of false memories differently depending on participants' prosociality. Specifically, participants with high prosociality scores reported more false memories in the selfish condition than in the social condition  $\beta = 0.68$ ,  $SE = 0.07$ ,  $t(42) = 2.42$ ,  $p = .02$ , 95% CI [0.03, 0.33], supporting Hess' findings on the negativity bias (Hess et al., 2013). This study provides an initial investigation into how social context shapes false memories. The results may suggest that negativity bias has a stronger influence on memory than the need to conform to prosocial beliefs (Bregman, 2020; Reicher & Haslam, 2006). To validate these findings, further research with larger samples and more effective manipulations is needed.

*Keywords:* false memories, pro-social/selfish context, misinformation paradigm, consistency, memory creation

## 1. Introduction

Although humans are often described as generally selfish, showing ingroup favouritism, self-interest and moral biases in social judgement and memory (Tajfel & Turner, 2001; Brewer, 1999), there is substantial evidence highlighting their motivation for cooperation, empathy and altruism. This underscores their nature as inherently social beings (Tomasello, 2014). This study aims to investigate whether observing prosocial and selfish behaviour in a social context influences memory organisation depending on the beliefs and personal social characteristics of the participant and if a prosocial context facilitates the formation of false memories, given the assumption that humans are inherently prosocial.

It has been shown that the valence (positive or negative) of social features modulates memory recollection. For instance, negative, social stimuli seem to be remembered better than positive, social stimuli, which is referred to a negativity bias (Hess et al., 2013). However, the negativity bias is context-dependent and is usually due to threatening content, like loss, grief or distress etc. as found in the material used in Hess' study (Hess et al. 2013). The authors found that the negativity bias were strongest when the emotional content of informative details was high, suggesting that threat-relevant social information enhances the memory of negative stimuli. This finding lends further support to the notion that negativity biases in memory are not automatic, but context dependent.

Importantly, the emotional valence of the stimuli is not the same dimension as the *social* valence of the stimuli, where research is still lacking. Hess and colleagues (2013) found that social stimuli activate different brain structures compared to nonsocial stimuli, particularly involving the medial prefrontal cortex, which is associated with self-referential processing. In a

second experiment participants were asked to imagine themselves or someone close to them in the situations depicted in the pictures (self-referential processing) and there was an enhancement in memory for positive social images. According to the self-perception theory, people often infer their own attitudes and preferences by observing their own behavior, much as they would infer the attitudes and preferences of others (Bem, 1967; Critcher & Gilovich, 2010). This theory states that people use their own behavior as a cue to understand their internal states and attitudes, and they may draw similar inferences when trying to understand the behavior of others. Harvey et al. (2007) propose that this might reflect the idea that social stimuli engage self-referential processes as individuals try to understand the mental states of others.

### **Congruency and coherence with own beliefs influence the formation of false memories**

The literature has documented that our memory is not always truthful and that it allows false memories to arise, which can have various causes; one of them being previous beliefs. In the context of generating false memories, individuals may incorporate details into their memories based on their beliefs and attitudes, leading to the creation of distorted or inaccurate recollections. According to this Frenda et al. (2013) found that false memories are implanted and recalled more easily when the false information is congruent with one's attitudes and beliefs, at least in a political context (Frenda et al., 2013; Armaly & Enders, 2023). Additionally, Zuh and colleagues (2010) mention that memories are not always accurate representations of past experiences but are influenced by personal meanings and interpretations. The illusory truth effect is another indicator for the assumption that congruent behavior makes false memories more likely, since the truth effect arises because people use peripheral cues such as familiarity (Begg et al., 1992), fluency, cohesion or congruence as cues for judging truth. Additionally, Conway and colleagues (2004) demonstrated that individuals tend to prefer coherence and

cohesion in memories by showing that people organise autobiographical memories so that they correspond to their current self-image and life story, even if this requires the falsification or omission of details in order to maintain internal consistency and psychological stability.

Therefore, we could assume that false information in general might be accepted and integrated in one's memory more easily when the context is congruent with the participants beliefs and attitudes in general.

### **Prosocial behavior as a congruent condition**

Contrary to much research that emphasize the selfish or aggressive aspects of human nature (e.g. Zimbardo, 1971; Milgram, 1963), Bregman (2020) and other authors present a vast amount of research and public events that show that humans are inherently prosocial, cooperative and are guided by empathy under typical conditions (Bregman, 2020, pp. 20,21). Building on this view, we hypothesize that human nature is generally pro-social and therefore a pro-social context would provide a congruent situation for participants, leading to a facilitation of false memory creation.

While Richard Dawkins' metaphor of the "selfish gene" popularized the idea that evolution is driven by individual genetic competition (Dawkins, 1976), subsequent biological research and reinterpretations argue for a fundamentally more cooperative framework. As described in Quanta Magazine, genes do not act in isolation but function as an interdependent network that produces the complex mechanisms of life through teamwork rather than conflict. Mark Ridley (2000) argues in *The Cooperative Gene* that genetic success depends on the ability of genes to cooperate, similar to players on a sports team or scientists in a complex project. This interpretation is shared by evolutionary theorists such as Axelrod and Hamilton (1981), who showed that cooperation can arise and persist even in competitive environments through repeated

interactions, and by Nowak et al., (2010), who showed that complex social behaviors can evolve through group and kin selection.

Furthermore, ethologist Frans de Waal (2013) supports the idea that morality, empathy, and altruism are not cultural artifacts, but biologically developed traits with deep evolutionary roots. Although isolated cases of “selfish” genetic behavior, such as transposons or meiotic drive elements, have been documented (Burt & Trivers, 2006), these are the exception rather than the rule. In most cases, genes thrive not by undermining others, but by contributing to the survival and reproductive success of the organism as a whole. This cooperative architecture at the genetic level reinforces the idea that humans are by no means fundamentally selfish, but are biologically equipped for prosocial behavior.

Prosocial behavior is not only visible when understanding our genetic evolution, as Kahneman and colleagues (1986) demonstrated that community norms of fairness significantly limit profit-seeking behavior in the markets. Their results show that 82% of the participants found it unfair that a hardware store raises the price of snow shovels from \$15 to \$20 the morning after a large snowstorm, and 77% perceived it as unfair when a company experiencing severe unemployment in the area and decides to reduce current workers' wages by 5%. This preference indicates a concern for equal treatment and social well-being, highlighting an intrinsic motivation to ensure fair outcomes for others and to promote trust and collective well-being by discouraging exploitative practices that harm others. It runs in contrast to conservative economic, which assume that individuals and companies would mainly act out of rational self-interest to maximize their own benefits. Strengthening this perspective, Zimbardo et al. (1971) showed in his Stanford Prison experiment, how aggressive, selfish and immoral people are when feeling powerful in a role and civilization with its rights, values and moral standards is taken away from

them. However, it was only decades later that the scientific community recognized that the results of the experiment were likely influenced by Zimbardo's own beliefs and expectations (Reicher & Haslam, 2006). Far less well known is a replication of this study conducted by the BBC, which produced opposite results. Here, researchers did not threaten and manipulate their participants, resulting in pro-social behavior of both prisoners and guards. They did not take their roles very seriously and instead started chatting and playing games with each other (Reicher & Haslam, 2006). To conclude, although researchers often emphasise the negative aspects of human nature, such as selfishness and aggression, in order to explain apparent cases of antisocial behaviour, prosocial tendencies appear to be intrinsic to the majority of people. Given this, and assuming that congruency with prior beliefs and expectations facilitates the formation of false memory, we considered a prosocial context as more likely to shape memory processes and increase the likelihood of false memory formation than a selfish context.

### **The paradigm of false memories**

In 2015 Julia Shaw and Stephen Porter convinced 70% of their participants that they had committed a crime in their youth, by completing three (40 min long) interviews using a basic interview script (Shaw & Porter, 2015). This sparked a global debate about the validity of human memories, the validity of Shaw's study and possible implications. It has been strongly criticized that Shaw did not distinguish between false memories and false beliefs in her experiment, although such distinction has been considered crucial in previous research (Wade et al., 2018). In 2018, Wade et al. replicated Shaw's experiment with a different evaluation scheme and found 26-30% false memories and around 40% false beliefs in their sample (Wade et al., 2018). These results are consistent with previous studies, such as the "lost in the mall" paradigm, where 29.2% of participants falsely recalled being lost in a shopping mall at the age of five (Loftus &

Pickrell, 1995). However, Shaw and other scientists see a fundamental problem in the construct of memory and the differentiation between belief and memory. According to Shaw, memories are beliefs about what happened and, conversely, beliefs are constructed and influenced by memories (Shaw, 2018). Following Shaw, a false memory is subject to some specific criteria, such as the agreement of a false event with the mention of at least ten false details (Shaw & Porter, 2015). In addition, critical details must be remembered, such as the specific location or the specific person who was allegedly present. In comparison, according to Hyman and Billings (1998), false memory only occurs when reports contain critical misinformation that has been previously dispersed (such as having spilled a punch) and when the processing and details in reports are consistent (Hyman & Billings, 1998). Thus, a major problem in studies of false memories is that definitions are often missing or variable (cf. Hyman & Billings, 1998; Shaw & Porter, 2015). In the present study, a false memory arises from previously disseminated misinformation within the setup and can therefore be assigned to the misinformation paradigm (Loftus et al., 1978).

Loftus found that there are factors that make the appearance of a false memory more likely or less likely, such as suggestion or pictorial imagination (Loftus, 1998). Misinformation in a conversation or the pictorial representation of a situation (effect of imagination inflation) make it easier to create false memories (Garry et al., 1996). The post event misinformation effect described by Loftus and colleagues in 1978 can be used as a tool to create false memories (Loftus et al., 1978). People witness a virtual scene (such as a robbery), are then exposed to misinformation and are finally asked to remember as many details as possible. Here a false memory arises from a vivid, experienced event and is therefore likely to create a mental, moving image, which results in a high probability of integration and retrieval from episodic memory, and

thus comes close to real memories (Kosslyn et al., 1979). Later findings of Phelps and Sharot (2008) support this hypothesis by investigating the influence of emotions on the subjective experience of memory, observing neurological mechanisms (Phelps & Sharot, 2008). Utilising functional magnetic resonance imaging (fMRI) and memory tasks with emotionally arousing and neutral stimuli, they found that memories associated with strong emotions were recalled more vividly and confidently, even in the absence of enhanced accuracy. They concluded that emotional content enhances subjective memory through interactions between the amygdala and hippocampus, rendering such memories more susceptible to distortion when suggestive or misleading information matches the emotional tone (Phelps & Sharot, 2008). Apart from emotional content it has been shown that social pressure and normative influence cause individuals to report false memories for the purpose of conforming to a group consensus. This phenomenon is known as memory conformity (Wright et al., 2000). These findings imply that false memories are not exclusively the result of cognitive errors or external suggestions, but can also be a product of social construction, especially when misinformation aligns with prior beliefs or socially desirable interpretations.

### **Overview of the study**

Using the misinformation paradigm, we created a pro-social and a selfish precondition where participants were presented with a short text about the social situation of a guy (target) they saw later in a slideshow breaking into a car. Using the common setup in misinformation experiments, participants had to recall the event later, after being exposed to misinformation. Since false memories increase with congruence (Frenda et al., 2013; Armaly & Enders, 2023), we expect participants to generate more false memories for the context that better matches their own beliefs about human nature. Therefore, a prosocial tendencies measure (Carlo & Randall,

2002) was integrated into this experiment, in order to measure the social beliefs and tendencies of each participant and to analyze its potential moderating role in the formation of false memories. Additionally, we assume that those exposed to a prosocial precondition, which seems to be congruent to the majority of humans, to generate more false memories in the recall test than those in the selfish precondition. It is crucial to note that a comprehensive validation of our hypothesis can only be achieved if the manipulation effectively impacts the participants' perception of the target individual in the slideshow. Specifically, if they believed that the target (guy) broke into the car and stole items for prosocial reasons, such as to pay for treatment for his sick daughter, or if they believe that the man was acting solely out of self-interest. The participants' perceptions of the target person (manipulation check) were therefore evaluated and analysed as a moderating variable to investigate how these perceptions might have influenced the effect of the experimental manipulation on the formation of false memories.

## **2. Methods**

### **2.1 Participants**

Voluntary subjects were randomly recruited via social media platforms and private distributions (mean age = 35,4 years; 66 % female). Since there is no previous research about the social valence of false memories, regarding pro-social or selfish behavior no expected effect can be assumed. Therefore, a mean Cohen effect of  $d = 0.25$  for F values was assumed. The sample size was calculated using G\*Power version 3.1.9.6 (Faul, Erdfelder, Lang & Buchner, 2007). With a type I error probability of  $\alpha = .05$ , a test power of  $(1-\beta) = .8$ , a total sample size of 128 subjects would have been required (“ANOVA: Fixed effects, omnibus, one-way”). In this study, 63 participants began the online experiment, 45 of whom completed it successfully, which

is sufficient to identify trends (outliers excluded). Participants under the age of 18 were excluded from the study and who took longer than 2 hours to complete the experiment.

## **2.2 Design**

The present study adopts a quantitative, experimental approach with a design that contrasts two groups (prosocial and selfish context) in an online setting, using a computer simulation (slideshow + 50 phrases which contain misinformation) followed by a recall test (Okadao & Stark, 2005). Accordingly, this design involved the comparison of two conditions on a single, continuous dependent factor (the sum of false memories), employing a between-subjects experimental design. Participants were randomly assigned to one of two conditions, and the total number of falsely recalled details served as the outcome measure. Furthermore, individual differences in pro social tendencies and subjective perceptions of the target were assessed as potential moderators to be included in subsequent analyses (see Material & Instruments and Result sections).

## **2.3 Material & Instruments**

The slideshows were developed by Okado & Stark (2005) and later adapted by Zhu et al. (2010). This experimental design was used for the present study, with only a social vs. selfish pre-coding phase in the beginning. Therefore, a pre-test was conducted (see Stimulus Development). Subjects are shown one event, which was depicted in a sequence of 50 digital color slides (Okado & Stark, 2005; Zuh et al., 2010). The event featured a man breaking into a car and stealing items and money. The methods used in this study have been supported by previous research (Okado & Stark, 2005; Zuh et al., 2010). These studies found consistent results, lending further credibility to the methods' validity and reliability. A manipulation check was conducted in order to ensure that the man's behavior in the slideshow was indeed perceived

as either pro social or selfish. Therefore, the same person perception scale (Fiske, Cuddy & Glick, 2007) as in the pretest was used, containing 14 adjectives to describe the perceived man's personality, ranging from empathic to selfish (see Stimulus development). In addition, a personality measure assessing the prosocial tendencies and beliefs of each participant was included in order to examine its potential moderating role in the formation of, or resistance to, false memories. Therefore the Prosocial Tendencies Measure (PTM), developed by Carlo and Randall (2002), was used. This scale contains 23-items and is based on the theoretical distinction between different types of prosocial disposition, measuring six sub-dimensions (e.g. public, anonymous, direct, emotional, indulgent and altruistic prosocial tendencies). These sub-dimensions are often analysed in terms of overarching factors such as prosocial motivation, social responsibility and empathy-driven helping (Rodrigues et al., 2017). The two scales were simultaneously used as filler tasks between the encoding phase and the misinformation phase, following the procedure from previous studies (Okado & Stark, 2005; Zuh et al., 2010). Between the misinformation phase and the recall test another filler task, specifically four items from the Cognitive Reflection Test (CRT) (Frederick, 2005) were integrated (Okado & Stark, 2005; Zuh et al., 2010)<sup>1</sup>.

---

<sup>1</sup> To effectively interrupt memory consolidation within a false memory experiment, it is crucial that participants are sufficiently challenged by filler tasks to ensure the necessary distraction from the encoded material. While previous studies often use time-based filler tasks of a longer duration (Zuh et al., 2010), this study employed short, cognitively demanding tasks, such as the four items from the Cognitive Reflection Test (CRT) (Frederick, 2005). These tasks, e.g. the classic bat-and-ball problem or the lily-pad scenario, require the inhibition of intuitive but incorrect responses and therefore address executive functions such as cognitive control and analytical thinking (Toplak et al., 2011). According to working memory theory, this system has a limited capacity and can become overloaded when high-level reasoning is required (Baddeley & Hitch, 1974). This is supported by findings showing that short, demanding cognitive tasks can impair memory performance as effectively as longer filler tasks by interrupting consolidation (Dewar et al., 2012), which is why we used these tasks in this experiment.

## Stimulus Development

To adapt the experimental setup to include a social and a selfish condition, we introduced a pre-coding text before participants were exposed to the slideshow. For the social condition, the pre-coding text provided context about the man's situation, explaining that he has a daughter with a rare disease who requires expensive treatment that he cannot afford (...). Conversely, for the selfish condition, the pre-coding text described the same family and financial situation, adding the man's desire for a luxurious lifestyle- cruise ships, designer clothes and champagne (see attachment). To validate the credibility of the Pre-coding texts a pretest was conducted in which participants rated the texts in terms of their perception of social and selfish behavior. Therefore, 40 participants were randomly assigned to either the social text or the selfish text and later asked to rate the man's personality on a person perception scale (Fiske et al., 2007), ranging from 1 (not at all) to 7 (very much). This scale captured the two dimensions empathy and selfishness with a total of 14 adjectives (e.g. 'selfish', 'self-serving', 'self-centred' vs. 'altruistic', 'kind', 'compassionate'). Although this version of the scale has not been officially published, it is theoretically based on the well-established 'Big Two' model of person perception (Fiske et al., 2007), which identifies warmth (also referred to as connectedness or morality) and competence (agency) as the two basic dimensions by which people evaluate others. The items in this manipulation check are particularly grouped around the warmth/community dimension, which includes traits such as empathy, benevolence, and social engagement as opposed to traits indicating selfishness or social distancing. These dimensions have been shown to influence impression formation, trust evaluations, and moral judgements (Abele & Wojciszke, 2014; Cuddy et al., 2008). Therefore, the use of trait-based ratings is supported by literature on person

perception, where trait adjectives serve as reliable indicators of moral and social character (Goodwin et al., 2014).

## **2.4 Procedure**

Each subject undergoes the same procedure. They were told that they are taking part in a study on the perception of visual stimuli. The study starts with demographic questions about the age and gender and an acknowledgement with the opportunity to contact the researchers. The first half of the experiment contained the pretext (either social or selfish), followed by the slideshow (Okadao & Stark, 2005), a manipulation check and a personality test concerning the pro social behavior of each participant, which were both simultaneously used as a filler task. The second half of the experiment contained the misinformation phase (50 sentences), a filler task and the final recall test.

After agreeing on the conditions of this study, participants were randomly assigned to one of two pre-coding texts (social vs. selfish). The slideshow contained 50 images, each of them was being displayed for 3,500 milliseconds with a 500-millisecond interval between slides. After viewing the slides, subjects completed 7 minutes of distraction tasks. These filler tasks contained the manipulation check (person perception scale) (Fiske et al., 2007) and a personality test (Frederick, 2005), which was later analyzed as a moderator variable. They then read a narrative about the event they have previously seen. The narrative consisted of 50 sentences, including 12 that provided misinformation and 38 that accurately described the scenes. The instructions didn't provide warnings of the misinformation in the narratives. Each sentence was shown for 3,500 milliseconds with a 500-millisecond interval between sentences, presented in the same order as the pictures to maintain the storyline. Following another filler task of 2-10 minutes (depending on the person) subjects took a three-alternative forced-choice recognition test (Original Event

item, Misinformation item, or Foil item). This test included 12 critical questions (out of 18) for the event, based on the slides they had seen and the narratives they had read. The questions are presented in random order, with three possible answers for each: the detail from the slide (original item), the misinformation detail from the narrative (misinformation item), or a new, unrelated detail (foil item). For example, if the slides showed a man hiding behind a door after stealing the girl's wallet, but the narrative stated he was hiding behind a tree, the critical question might ask, "Where was the man hiding after stealing the girl's wallet?" with the choices "behind the tree" (misinformation item), "behind the door" (original item), and "behind the car" (foil item). The foil items serve as control items. The study lasted between 12.2 and 58.8 minutes, as there were two consolidation (filler) times of around 7-15 and around 2-15 minutes, depending on the speed of each participant.

## **Results**

The IBM SPSS STATISTICS software program (Version 26.0.0.0; IBM, 2019) will be used for coding and statistical analysis of the data, as well as the software Jamovi. For the analysis of the present study the rates at which subjects affirmed the misinformation, the original and the foil items are used to calculate the false memory of the misinformation vs. the true memory. The internal consistency of the test assessing the construct of false memories was evaluated using Cronbach's alpha and is conducted to be  $\alpha = .05$ .

### *Analysis of psychometric properties of scales*

To further assess the psychometric properties of the person perception scale (Fiske et al., 2007) a confirmatory factor analysis (CFA) was conducted to test whether the data supported the hypothesized two-factor structure. The aim was to assess whether participants' perceptions of the

conditions were consistent with the respective constructs of the scale- selfishness and sociality, each with seven observed indicators. Model fit was suboptimal,  $\chi^2(76) = 140, p < .001$ ; CFI = 0.873; TLI = 0.849; RMSEA = 0.137, 90% CI [0.101, 0.172], indicating that while the factorial structure is supported, the overall model may benefit from further refinement. All items loaded significantly on their respective factors (all  $p < .05$ ), supporting the intended structure. Factor loadings for the social items ranged from 0.81 to 1.27, while those for selfish items ranged from .77 to 1.7. The latent factors were significantly negatively correlated ( $r = -.33, p = .02$ ), consistent with the theoretical assumption of opposing constructs.

A confirmatory factor analysis (CFA) was conducted to assess the latent structure of the Prosocial Tendencies Measure (PTM) (Carlo & Randall, 2002) and its six constructs: Altruistic, Public, Emotional, Dire, Compliant, and Anonymous prosocial behavior. The hypothesized model demonstrated poor fit,  $\chi^2(194) = 316, p < .001$ , RMSEA = .117, 90% CI [.093, .140], CFI = .664, TLI = .600. While the Emotional and Anonymous factors were well-defined with significant loadings across indicators (all  $ps < .001$ ), other factors (e.g. Public and Altruistic) showed weak or inconsistent loadings. Strong correlations were observed between Emotional and Dire ( $r = .87, p < .001$ ), and Dire and Compliant ( $r = .72, p < .001$ ), suggesting potential overlap. The overall findings indicate that the proposed model may require revision to better represent the data structure. Therefore, an exploratory factor analysis was conducted in a second step to test better suited dimensions (see below).

Exploratory factor analysis (EFA) - Due to the poor model fit observed in the confirmatory factor analysis (CFA) of the Prosocial Personality Scale, an exploratory factor analysis (EFA) was conducted using the method of minimum residual extraction with oblimin rotation. The Kaiser criterion (eigenvalues  $> 1$ ) was applied to extract factors from the 23 items in order to

identify a more suitable dimensional structure for the scale. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .60, indicating a moderate suitability for factor analysis. Bartlett's test of sphericity was significant,  $\chi^2(253) = 505$ ,  $p < .001$ , indicating sufficient correlations among variables for factor extraction. Communalities ranged from 0.291 to 0.907. A four-factor solution was extracted, with an acceptable model fit: RMSEA = .053, 90% CI [0.00, 0.092]. Factor loadings revealed a clear and interpretable structure across the four extracted dimensions. **Factor 1** consisted predominantly of items reflecting anonymous helping, with strong loadings from items 8, 11, 15, and 19 ( $> 0.604$ ) (see annex). **Factor 2** captured a mix of compliant, emotional, and crisis-oriented (dire) helping behaviors, indicated by substantial loadings from items 2, 6, 7, and 18 ( $> 0.55$ ). **Factor 3** was defined by high loadings on items related to altruistic helping, including items 16, 22, and 23 ( $> 0.543$ ). Finally, **Factor 4** represented public or status-driven helping, with strong contributions from items 1, 3, 5, and 4 ( $> 0.477$ ). Communalities ranged from moderate to high, with the majority of items exceeding .40, indicating good shared variance among extracted factors. A few items showed cross-loadings (e.g., items 21 and 17), which may reflect conceptual overlap between helping styles. Overall, the EFA supports a multidimensional structure of prosociality, confirming the distinct but correlated nature of the helping styles described in the original scale.

The four extracted factors were labeled based on their thematic content and item composition (see Table 1).

**Table 1.**

<b>Factor</b>	<b>Label</b>	<b>Main Theme &amp; Items</b>	<b>Example of Representative Items</b>
<b>1</b>	Anonymous and emotionally responsive helping	Helping anonymously, triggered by emotional or crisis-related situations Items: 19,11,15,8,21,17,14,10	“I prefer to donate money anonymously.” “Emotional situations make me want to help others.”
<b>2</b>	Compliant and crisis-oriented helping	Immediate helping in response to requests or emergencies Items: 7,18,2,6,9,12	“When people ask me to help them, I don’t hesitate.” “I help people in real crisis situations.”
<b>3</b>	Reciprocal and recognition-seeking helping	Helping motivated by personal benefit, reciprocity, or a need for acknowledgment Items: 16, 23, 22	“I feel that if I help someone, they should help me in the future.” “I should receive more recognition for charity work.”
<b>4</b>	Public and status-driven helping	Helping in visible, public settings; motivated by image or social rewards Items: 4,5,3,1,20,13	“Helping others when I am in the spotlight is when I work best.” “Charity work looks good on my resume.”

### *Manipulation checks*

To ensure that participants perceived the man’s behavior in the video as selfish or social, depending on the pretext they have seen, we conducted an independent sample t-test to compare the mean ratings of the social items and the selfish items from the person perception scale (Fiske et al., 2007), between the two conditions (social vs. selfish pretext).

We first analyze the total score for the selfish items revealed a significant difference between conditions,  $t(43) = -3.66, p < .001, d = 1.1$ , with participants in the selfish condition

reporting higher endorsement of selfish behaviors ( $M= 5.56$ ,  $SD= .99$ ) compared to those in the social condition ( $M= 4.08$ ,  $SD= 1.69$ ), see tabel 2 below.

We then analyzed item by item. All individual items also showed significant differences in the expected direction (all  $p < .014$ ), with large effect sizes (Cohen's  $d$  ranging from  $-0.77$  to  $-1.01$ ), indicating a robust effect of the manipulation on selfish tendencies. Participants indeed perceived the man's behavior as more selfish when seeing the selfish pretext, compared to the social pretext. In contrast, no significant differences were found for the social items depending on the condition (social condition:  $M = 2.32$ ,  $SD = 1.14$ ; selfish condition  $M = 2.54$ ,  $SD = 1.12$ ), either for the total score,  $t(43) = -0.65$ ,  $p = .52$ ,  $d= -0.2$ , nor for any individual item (all  $p > .39$ ).

**Table 2.**

Selfish items	Condition	M	SD	Social Items	M	SD
2 selfish	Social	4.4	1.93	1 empathic	2.4	1.43
	Selfish	5.8	1.16		2.6	1.47
4 egoistic	Social	4.35	2.06	3 friendly	2.8	1.36
	Selfish	5.72	1.28		3.2	1.66
5 self- serving	Social	4.2	2.33	10 altruistic	2.15	1.27
	Selfish	5.72	1.65		2.32	1.25
6 egocentric	Social	3.7	2.11	11 generous	1.9	1.02
	Selfish	5.4	1.26		2.2	1.44
7 self-absorbed	Social	3.75	2.12	12 compassionate	2.3	1.26
	Selfish	5.16	1.34		2.48	1.29
8 self- interested	Social	4.4	2.3	13 benevolent	2.25	1.33
	Selfish	5.84	1.18		2.52	1.53

9 uncaring	Social	3.75	2.17	14 kind-hearted	2.45	1.54
	Selfish	5.28	1.4		2.48	1.33
SELFISH ALL	Social	4.08	1.69	SOCIAL ALL	2.32	1.14
	Selfish	5.56	1.0		2.54	1.12

### General accuracy and false memories

The mean rate of endorsement of the “original items” was 64,93 %, indicating a high level of accuracy of memory. The mean endorsement rate for misinformation items was 20,37 %, and for foil items (misses) 14,69 %. This rate of false memory inducement is consistent with previous research (Okado & Stark, 2005; Zhu et al., 2010). Additionally, Pearson correlations showed a strong negative correlation between false and true items,  $r(43) = -.872, p < .001$ , and between true and foil items,  $r(43) = -.564, p < .001$ . Foil and False items showed a non-significant correlation of  $r(43) = .087, p = .568$ , supporting previous studies (Zuh et al., 2010; Okado & Stark, 2005) and lending further credibility to this study. A paired-samples t-test revealed that the endorsement rate for misinformation items was significantly higher than for foil items,  $t(44) = 2.04, p = .047$ , Cohen’s  $d = 0.30$ , indicating a small to moderate effect. These results indicate that participants were more likely to endorse misinformation items than foil items, supporting the idea that the paradigm successfully elicited false memories,

### Hypothesis: 1 Prosocial vs. Selfish context impacts the creation of False Memories

Contrary to our expectation, we find no effect on the formation of false memories regarding the two groups (social:  $M = 3.52, SD = 2.84$ , vs. selfish:  $M = 3.72, SD = 3.16$ ),  $t(44) = -.220, p = .827$ , with a negligible effect size (Cohen's  $d = -.065$ ). These findings suggest that the context, whether selfish or social, did not influence the occurrence of false memories in general.

### Hypothesis 2: Prosocial personality traits and context impact false memory creation

To examine whether the effect of context (social vs. selfish) on false memory formation is moderated by individual differences in prosocial personality traits, we first considered the total score of the Prosocial Tendencies Measure (PTM) (Carlo & Randall, 2002), followed by separate analyses of each personality trait. A series of General Linear Models (GLMs) was performed for this purpose.

Total score- A first model tested the interaction between condition and the total prosociality score and revealed a significant interaction,  $F(1, 42) = 5.87, p = .02, \eta^2_p = .12$ , indicating that the context influenced false memory formation differently depending on participants' level of prosociality. Specifically, participants with high prosociality scores reported more false memories in the selfish condition than in the social condition  $\beta = 0.68, SE = 0.07, t(42) = 2.42, p = .02, 95\% CI [0.03, 0.33]$ . No significant main effects of condition or prosociality was found ( $ps > .089$ ).

Altruistic Dimension- A separate model, focused on the interaction between altruistic traits and condition, revealed a significant interaction effect,  $F(1, 42) = 5.05, p = .03, \eta^2_p = .107$ . Post hoc analyses clarified the nature of the interaction. Participants who scored high on the altruistic dimension reported significantly more false memories in the selfish condition compared to the social condition,  $\beta = 0.66, SE = 0.36, t(42) = 2.33, p = .03, 95\% CI [0.08, 1.51]$  Neither the main effect of condition nor altruism alone was significant ( $ps > .82$ ).

Emotional Dimension- A second model assessed the emotional dimension of prosocial personality and found a significant interaction effect between emotional prosocial traits and condition,  $F(1, 42) = 5.21, p = .028, \eta^2_p = .11$ . According the trends we observed previously, participants who scored high on the emotional dimension reported significantly more false memories in the selfish condition compared to the social condition,  $\beta = 0.65, SE = 0.23, t(42) =$

2.28,  $p = .028$ , 95% CI [0.06, 0.96]. There were no significant main effects of condition ( $p = .96$ ) or emotional traits ( $p = .09$ ).

Dire Dimension - Another model assessed the dire dimension of prosocial personality and found a significant interaction effect,  $F(1, 42) = 4.54$ ,  $p = .039$ ,  $\eta^2_p = .097$ . In line with previously observed trends, participants who scored high on the dire helping dimension reported significantly more false memories in the selfish condition compared to the social condition,  $\beta = 0.63$ ,  $SE = 0.35$ ,  $t(42) = 2.13$ ,  $p = .039$ , 95% CI [0.04, 1.43]. There were no significant main effects of condition ( $p = .82$ ) or dire helping traits ( $p = .92$ ).

As the original dimensions of the Prosocial Tendencies Measure (PTM; Carlo & Randall, 2002) could not be confirmed through confirmatory factor analysis (CFA), we examined whether the newly identified dimensions from the exploratory factor analysis (EFA) moderated the effect of context (prosocial vs. selfish) on false memory creation. To test this, four separate general linear models were conducted, with each model testing one of the extracted EFA factors as a moderator variable.

1. Anonymous and emotionally responsive helping - The overall model was not significant,  $F(3, 42) = 1.62$ ,  $p = .198$ ,  $\eta^2_p = .104$ . There were no significant main effects of condition,  $F(1, 42) = 0.01$ ,  $p = .937$ , or anonymous and emotionally responsive helping,  $F(1, 42) = 3.44$ ,  $p = .071$ . The interaction between condition and anonymous and emotionally responsive helping was also not significant,  $F(1, 42) = 1.12$ ,  $p = .296$ .

2. Compliant and Crisis-Oriented Helping - This model approached significance overall,  $F(3, 42) = 2.35$ ,  $p = .086$ ,  $\eta^2_p = .144$ . There was a significant interaction between condition and compliant and crisis-oriented helping,  $F(1, 42) = 6.12$ ,  $p = .017$ ,  $\eta^2_p = .127$ .

Participants who scored high on the compliant and crisis-oriented helping dimension reported significantly more false memories in the selfish condition compared to the social condition,  $\beta = 0.72$ ,  $SE = 0.18$ ,  $t(42) = 2.47$ ,  $p = .017$ , 95% CI [0.08, 0.83].

Main effects for condition ( $p = .922$ ) and compliant and crisis-oriented helping itself ( $p = .199$ ) were not significant.

3. Reciprocal and Recognition-Seeking Helping - This model was not significant,  $F(3, 42) = 0.78$ ,  $p = .514$ ,  $\eta^2_p = .052$ . There were no significant main effects of condition,  $F(1, 42) = 0.05$ ,  $p = .825$ , or reciprocal and recognition-seeking helping,  $F(1, 42) = 0.36$ ,  $p = .551$ . The interaction also did not reach significance,  $F(1, 42) = 2.23$ ,  $p = .143$ .

4. Public and status-driven helping - The model approached significance,  $F(3, 42) = 2.44$ ,  $p = .077$ ,  $\eta^2_p = .149$ . Importantly, the interaction between condition and public and status-driven helping was significant,  $F(1, 42) = 4.31$ ,  $p = .044$ ,  $\eta^2_p = .093$ . According to previous trends, participants who scored high on the public and status-driven helping dimension reported significantly more false memories in the selfish condition compared to the social condition,  $\beta = 0.69$ ,  $SE = 0.43$ ,  $t(42) = 2.07$ ,  $p = .044$ , 95% CI [0.03, 1.77]. Main effects of condition ( $p = .843$ ) and the prosocial factor ( $p = .702$ ) were not significant.

*Hypothesis 3: Context and prosocial personality impacts creation of false memories only for those who perceived the target congruently with the manipulation*

We hypothesized that some participants might not be sensitive to our manipulations, which could prevent us from properly testing our hypothesis. Therefore, it was important to determine whether the effects were observed at least in these participants. As a result, only

participants in whom the experimental manipulation was successful were included in this analysis. Manipulation success was defined based on participants' ratings of the target (guy) on the 14 trait adjectives of the person perception scale (Fiske et al., 2007). A difference score was calculated by subtracting the mean rating of selfish traits from the mean rating of social traits. If the difference was positive in the social condition or negative in the selfish condition, the participant was classified as having successfully perceived the manipulation. To examine whether the effect of the contextual manipulation (social vs. selfish motive) on false memory formation was moderated by individual differences in prosocial tendencies among participants who successfully perceived the manipulation, several General Linear Models (GLMs) were conducted. These analyses focused, again, first on the total score of the scale and then on each personality trait individually.

A first model tested the interaction between experimental condition and overall prosociality (total score). Although the interaction did not reach conventional significance levels, it was marginally significant,  $F(1, 24) = 3.27, p = .083$ , partial  $\eta^2_p = .12$ . Participants high in prosociality tended to report more false memories in the selfish condition compared to the social condition,  $\beta = .749, SE = .1, t(24) = 1.81, p = .083, 95\% \text{ CI } [-0.027, .41]$ . To further explore which aspects of prosociality might drive this interaction, additional GLMs were calculated using each of the six prosocial subscales individually. None of the subscale models revealed a statistically significant interaction between condition and the respective trait (all  $ps > .11$ ).

## Discussion

The aim of this study was to identify differences in the formation of false memories depending on the social context of a situation. We hypothesized that participants who were given

a prosocial context about a man breaking into a car (to collect money for his sick daughter) would be more likely to form false memories, as this could represent a situation that is consistent with most people's beliefs and behaviors. We assumed that most people are more prosocial than selfish due to human nature, and therefore assumed that prosocial behavior is consistent with most people's beliefs. Results show no significant main effect of condition in the creation of false memories. However, an exploratory trend was observed, suggesting that individuals with higher levels of prosociality, measured by the Prosocial Tendencies Measure (Carlo & Randall, 2002), are more susceptible to false memories in the selfish condition than in the prosocial condition. Contrary to our hypothesis, participants with high prosociality scores reported more false memories in the selfish condition than in the social condition  $\beta = 0.68$ ,  $SE = 0.07$ ,  $t(42) = 2.42$ ,  $p = .02$ , 95% CI [0.03, 0.33], supporting Hess' findings on the negativity bias (Hess et al., 2013). This trend could be observed in several prosocial personality traits, such as altruistic, emotional, and dire behavior. The same was true in the new dimensions identified in the EFA: complaint & crisis-oriented help, as well as public & status-oriented help. Participants who scored high in those traits reported more false memories in the selfish condition than in the social condition. When reanalyzing only those who perceived the goal as congruent with manipulation, the same trend can be observed: Participants with high prosociality tended to report more false memories under selfish conditions than under social conditions,  $\beta = 0.749$ ,  $SE = 0.1$ ,  $t(24) = 1.81$ ,  $p = 0.083$ , 95% CI [-0.027, 0.41].

Although this result contradicts our original hypothesis, similar effects have been theorized in the literature. For example, schema theory assumes that when information received is incompatible with existing moral or social schemas, individuals may engage in cognitive reconstructions to restore coherence, which can lead to memory distortions. Similarly, research

on autobiographical memory shows that people tend to favor memories that are consistent with their identity (Conway & Pleydell-Pearce, 2000). This suggests that morally dissonant stimuli, such as selfish behavior observed in highly prosocial individuals, may challenge internal expectations and lead to greater susceptibility to false memories. This explanation would support a variety of studies on the prosocial nature of humans in general (e.g., Bergman, 2020). The finding that analogous trends were identified in both the original PTM subscales (altruistic, emotional, dire) and the newly derived EFA-based factors (e.g., compliant & crisis-oriented helping, public & status-oriented helping) tentatively supports the hypothesis that this potential effect is consistent.

Due to the small sample size, particularly within the subgroup that passed the manipulation check, all observed effects should be interpreted with caution and considered as preliminary tendencies rather than definitive conclusions. The general absence of a significant main effect of context (prosocial vs. selfish) on false memory formation may reflect a lack of statistical power rather than a genuine lack of effect (Maxwell, 2004). As Simmons et al. (2011) emphasize, the use of small samples can significantly increase the likelihood of first- and second-type errors.

Another limitation of the present study concerns the dimensional structure of the Prosocial Tendencies Measure (PTM; Carlo & Randall, 2002). Despite the widespread use of the PTM as an instrument for assessing various forms of prosocial behavior, the present study suggests that its dimensional structure cannot be well replicated in this particular sample. In particular, confirmatory factor analysis (CFA) failed to validate the original six-factor model, suggesting that some of the subdimensions may not be clearly distinguishable or psychometrically robust in this context. This phenomenon may be attributed to conceptual overlap between certain

subscales or the situational dependence of prosocial motivations, as has been highlighted in critiques of multidimensional measures of prosociality (e.g., Padilla-Walker & Carlo, 2014). Therefore, it is recommended that future studies use an alternative personality instrument with clearer and theoretically grounded factor structures. Alternatively, research could focus more on the core subdimensions of prosociality, such as altruistic, emotional, and dire prosocial behavior, which showed more consistent patterns in the present study.

As for the manipulation itself, it should be noted that the misinformation disseminated and the subsequent memory test referred to neutral facts such as the color of a car, rather than prosocial versus selfish behavior. Further research could investigate possible effects when misinformation is disseminated in a prosocial vs. selfish direction (e.g., disseminating misinformation that the target person in the slideshow opened the door for a man leaving the store) rather than priming participants in one direction before they observe critical behavior by a fellow citizen. This could increase the success of manipulation among participants.

However, defining false memories in general remains a major challenge that goes beyond this study and affects the entire field of research. Even more controversial is the creation of false memories in studies, as this requires a clear definition of the construct, which does not exist. Furthermore, the replication of an actual false memory is almost impossible. In this study, false memories were suggestively created through misinformation within the setting. The subject's personal involvement in the experience is relatively low, which raises the question of whether simply witnessing the slideshow is sufficient to activate episodic memory, or whether the memories are merely based on familiarity phenomena (similar to DRM false memories, where test subjects are asked to memorize words from word lists on a specific topic that are read aloud to them, and participants tend to mentally supplement these lists with matching, similar words

and then remember them (cf. DRM effect). To ensure that the memory is retrieved from episodic memory and not from working memory (familiarity effect), future studies could use a similar experimental design but conduct the memory test after a longer period of time (e.g., two weeks), instead of using cognitively demanding filler tasks as in this study.

### References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions* (pp. 27–29). Sage.
- Armaly, M. T., & Enders, A. M. (2023). Filling in the gaps: false memories and partisan bias. *Political Psychology, 44*(2), 281-299.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47–89. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Begg, I. M., Anas, A. & Farinacci, S. (1992). Dissociation of processes in belief: source recollection, statement familiarity, and the illusion of truth. *J. Exp. Psychology Gen. 121*, 446– 458.
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological review, 74*(3), 183.

- Bernstein, D. M., Scoboria, A., Desjarlais, L., & Soucie, K. (2018). "False memory" is a linguistic convenience. *Psychology of Consciousness (Washington, D.C.)*, 5(2), 161–179.  
<https://doi.org/10.1037/cns0000148>
- Bregman, R. (2020). *Im Grunde gut. Eine neue Geschichte der Menschheit*, 4. Rowohlt Taschenbuch.
- Carlo, G., & Randall, B. A. (2002). The development of a measure of prosocial behaviors for late adolescents. *Journal of youth and adolescence*, 31, 31-44.  
<https://doi.org/10.1023/A:1014033032440>
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, 107(2), 261.
- Conway, M. A., Singer, J. A., & Tagini, A. (2004). The self and autobiographical memory: Correspondence and coherence. *Social cognition*, 22(5: Special issue), 491-529.
- Critcher, C. R., & Gilovich, T. (2010). Inferring attitudes from mindwandering. *Personality and Social Psychology Bulletin*, 36(9), 1255-1266.
- Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological science*, 23(9), 955-960.  
<https://doi.org/10.1177/09567976124412>

Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology, 49*(2), 280-286.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. <https://doi.org/10.1257/089533005775196732>

Galow, A., & Tamm, S. (2008). Erlebnisbasierte und verfälschte Erinnerungen im EEG: Eine Pilotstudie. *Forensische Psychiatrie Psychologie Kriminologie, 2*(1), 57–64.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*(2), 208–214. <https://doi.org/10.3758/BF03212420>

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on psychological science, 9*(6), 641-651.

Harvey, P. O., Fossati, P., & Lepage, M. (2007). Modulation of memory formation by stimulus content: Specific role of the medial prefrontal cortex in the successful encoding of social pictures. *Journal of Cognitive Neuroscience, 19*, 351-362.

<https://doi:10.1162/jocn.2007.19.2.351>

Hess, T. M., Popham, L. E., Dennis, P. A., & Emery, L. (2013). Information content moderates positivity and negativity biases in memory. *Psychology and Aging*, 28(3), 853.

Hyman, I. E., Jr, & Billings, F. J. (1998). Individual differences and the creation of false childhood memories. *Memory (Hove, England)*, 6(1), 1–20.

<https://doi.org/10.1080/741941598>

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, (1986). “Fairness as a Constraint on Profit Seeking: Entitlements in the Market,” *American Economic Review*, 728–741.

Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). On the demystification of mental imagery. *The Behavioral and Brain Sciences*, 2(4), 535–548.

<https://doi.org/10.1017/s0140525x00064268>

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725. <https://doi.org/10.3928/0048-5713-19951201-07>

Loftus, E. F. (1998). Falsche Erinnerungen. *Spektrum der Wissenschaft*, 1, 62-67.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology. Human Learning and Memory*, 4(1), 19–31. <https://doi.org/10.1037//0278-7393.4.1.19>

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychological methods*, 9(2), 147.

<https://doi.org/10.1037/1082-989X.9.2.147>

Okado, Y., & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 12(1), 3–11.

<https://doi.org/10.1101/lm.87605>

Otgaar, H., & Candel, I. (2011). Children's false memories: different false memory paradigms reveal different results. *Psychology, Crime & Law: PC & L*, 17(6), 513–528.

<https://doi.org/10.1080/10683160903373392>

Padilla-Walker, L. M., & Carlo, G. (2014). The study of prosocial behavior: Past, present, and future. In L. M. Padilla-Walker & G. Carlo (Eds.), *Prosocial development: A multidimensional approach* (pp. 3–16). Oxford University

Press. <https://doi.org/10.1093/acprof:oso/9780199964772.003.0001>

Phelps, E. A., & Sharot, T. (2008). *How (and why) emotion enhances the subjective sense of recollection*. *Current Directions in Psychological Science*, 17(2), 147–

152. <https://doi.org/10.1111/j.1467-8721.2008.00565.x>

- Reicher, S., & Haslam, S. A. (2006). Rethinking the psychology of tyranny: The BBC prison study. *British Journal of Social Psychology*, 45(1), 1-40.
- Rodrigues, J., Ulrich, N., Mussel, P., & Carlo, G. (2017). Measuring prosocial tendencies in Germany: Sources of validity and reliability of the revised Prosocial Tendencies Measure. *Frontiers in Psychology*, 8, 2119. <https://doi.org/10.3389/fpsyg.2017.02119>
- Shaw, J., & Porter, S. (2015). Constructing rich false memories of committing crime. *Psychological Science*, 26(3), 291–301. <https://doi.org/10.1177/0956797614562862>
- Shaw, J. (2018). How can researchers tell whether someone has a false memory? Coding strategies in autobiographical false-memory research: A reply to Wade, Garry, and pezdek (2018). *Psychological Science*, 29(3), 477–480.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.  
<https://doi.org/10.1177/0956797611417632>
- Tajfel, H., & Turner, J. (2001). An integrative theory of intergroup conflict. In M. A. Hogg & D. Abrams (Eds.), *Intergroup relations: Essential readings* (pp. 94–109). Psychology Press.

- Tomasello, M. (2014). *A Natural History of Human Thinking*. Harvard University Press.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition*, 39(7), 1275-1289. <https://doi.org/10.3758/s13421-011-0104-1>
- Wade, K. A., Garry, M., & Pezdek, K. (2018). Deconstructing rich false memories of committing crime: Commentary on Shaw and Porter (2015). *Psychological Science*, 29(3), 471–476. <https://doi.org/10.1177/0956797617703667>
- Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology*, 91(2), 189–202. <https://doi.org/10.1348/000712600161781>
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... & Dong, Q. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences*, 48(8), 889-894.
- Zimbardo, P. G., Haney, C., Banks, W. C., & Jaffe, D. (1971). The Stanford prison experiment.

## Appendix 1

### Literature Review

#### **Introduction**

The idea that we have memories stored in our minds that never happened can sound frightening, surreal and can be hard to believe. The fact that we can convincingly describe such invented experiences, often without realizing that they are untrue, can be disturbing (Shaw & Porter, 2015). Julia Shaw and Stephen Porter's (2015) demonstrated this phenomenon by convincing 70% of participants in their sample that they had committed a crime in their youth (Shaw & Porter, 2015), by completing three (40 min long) interviews using a basic interview script. This sparked a global debate about the validity of human memories, the validity of Shaw's study and possible implications. In contrast, the "Lost in the Mall" paradigm, originally introduced by Loftus and Pickrell (1995) and frequently replicated ever since, used a much more harmless false memory and has shown significantly smaller effects. Although the event was very plausible and emotionally neutral, only 29.2% of participants falsely recalled being lost in a shopping mall at the age of five (Loftus & Pickrell, 1995). The "Lost in the Mall" paradigm is a fundamental experimental approach in the research of the formation of false memories. Similar to Shaw and Porter, a false narrative is embedded in real childhood memories and participants are encouraged to imagine the event. Using this method, participants described detailed and seemingly authentic memories of an event that had never actually taken place (getting lost in a Mall). It demonstrates how people can be led to believe in completely fabricated autobiographical events. This paradigm has played a significant role in memory research and forensic psychology, particularly in debates about the reliability of recovered memories in

therapeutic and legal contexts. However, recent critiques have questioned its ecological validity and ethical implications (Szojka et al., 2025). Critics argue that the paradigm overestimates the ease with which complex autobiographical false memories can arise, especially since the false event used (getting lost) is relatively plausible and emotionally neutral (Murphy & Greene, 2025). However, these controversial results, between the harmless false memory in the *Lost in the Mall* paradigm and the emotionally charged memories implanted in Shaw and Porter's study, sparked intense debate over whether Shaw's findings reflect the creation of genuine false memories or merely the formation of false beliefs (Wade et al., 2018). It has been strongly criticized that Shaw did not distinguish between false memories and false beliefs in her experiment, although such distinction has been considered crucial in previous research (Wade et al., 2018). So, whether someone remembers an event or only believes that something happened because there may be enough evidence and witnesses was not differentiated in Shaw's experiment, but it does represent a qualitative difference. In 2018, Wade et al. replicated Shaw's experiment with a different evaluation scheme and found 26-30 percent false memories and around 40 percent false beliefs in their sample (Wade et al., 2018). These results are consistent with previous studies, such as the "lost in the mall" paradigm (Loftus & Pickrell, 1995). However, Shaw and other scientists see a fundamental problem in the construct of memory and the differentiation between belief and memory. According to Shaw, memories are beliefs about what happened and, conversely, beliefs are constructed and influenced by memories (Shaw, 2018). In summary, she questions the purpose and usefulness of differentiation. For judicial context, this distinction seems less important, because whether a witness is *convinced* that he/she has seen e.g. a blue car or *remembers* a blue car, is of no interest to a judge in case of doubt. The statement "I saw a blue car" alone would usually be sufficient. The difference between a memory and a belief

is hardly distinguishable even for a person themselves after some time and especially after questioning, as we can see in the following sections.

### **The Concept of False Memories**

In the *Book of Mental Sciences*, a memory is described as the act of consciously retrieving something stored in the mind (Jordan, 2013). False memories, according to Shaw, must meet certain criteria, such as the agreement on a false event while recalling at least ten fabricated details (Shaw & Porter, 2015). Additionally, critical details such as the exact location or specific persons allegedly involved must be remembered.

By comparison, Hyman and Billings (1998) define false memories as those that include critical misinformation which have been spread earlier (e.g., spilling punch at a party) and ensure consistency in processing and reporting such details. The lack of standardized definitions in studies of false memories remains a significant issue in this field (Hyman & Billings, 1998; Shaw & Porter, 2015). In this study, a false memory arises from previously disseminated misinformation within the setup.

### **Influences on False Memories**

Due to its limited capacity, the human brain does not store exact copies of experiences, but instead encodes and retains essential information, leaving memories open to reconstruction and distortion (Schacter, 1999, 2001). Research on false memories identifies two key insights: there are factors that foster false memories and others that reduce their likelihood (Loftus, 1998). Suggestion and imagination are among the strongest contributors to the formation of false memories, as outlined by Elizabeth Loftus. Vividly imagining a scenario (called the imagination

inflation effect) make false memories more likely to emerge (Garry et al., 1996). In their experiment, participants were first asked to rate their certainty about whether certain events from their childhood (e.g., getting stuck in a tree or spilling punch at a wedding) had actually happened to them. Later, they were asked to imagine some of these events in detail. When questioned again, the participants showed significantly greater certainty that the imagined events had actually taken place in their past, even though there was no evidence to support this (Garry et al., 1996). Furthermore, Hyman and Billings (1998) showed that being confronted with misinformation in conversations, especially from trusted sources such as family members, can lead to the creation of completely false childhood memories. In their study, participants were presented with a mixture of real and fictional childhood events, and after being repeatedly questioned about the events, some “remembered” the fictional events as real.

Additionally, the further back an event lies in the past, the more susceptible it is to distortion. In experiments with misinformation, Loftus (1993, 2005) showed that participants were significantly more likely to integrate misleading details that were added after a longer period of time into their memory. This susceptibility of older, weaker memories explains why childhood memories, such as in the “Lost in the Mall” paradigm, are particularly prone to the formation of false memories (Loftus & Pickrell, 1995).

Nevertheless, the influences on the formation of false memories go beyond the experimental design, as even personality traits and current moods have an impact. Storbeck and Clore (2005) showed that people in a sad mood are less likely to develop false memories than those in a happy mood, as sad moods promote more detail-oriented processing and reduce reliance on essence-based associations. In addition, individuals who score high on personality

scales such as the Creation Imagination Scale (CIS) or the Dissociative Experiences Scale (DES) are more susceptible to false memories (Hyman & Billings, 1998). Results showed that traits such as a high tendency toward dissociation, a propensity for fantasizing, and creative imagination increase susceptibility to memory distortion, as these individuals are more likely to confuse internally generated thoughts or images with real experiences (Hyman & Billings, 1998).

### **Context dependency of memories**

It has been shown that negative, social stimuli seem to be remembered better than positive, social stimuli, which is referred to a negativity bias (Hess et al., 2013). However, the negativity bias seems to be context-dependent and is usually due to threatening content, like sadness, grief or distress etc. as found in the material used in Hess' study (Hess et al. 2013). The authors found that the negativity bias were strongest when the emotional content of informative details was high, suggesting that threat-relevant social information enhances the memory of negative stimuli. This finding lends further support to the notion that negativity biases in memory are not automatic, but context dependent. Nevertheless, the emotional valence of the stimuli is not the same dimension as the *social* valence of the stimuli, where research is still lacking. Hess and colleagues (2013) found that social stimuli activate different brain structures compared to nonsocial stimuli, particularly involving the medial prefrontal cortex, which is associated with self-referential processing. In a second experiment participants were asked to imagine themselves or someone close to them in the situations depicted in the pictures (self-referential processing) and there was an enhancement in memory for positive social images. According to the self-perception theory, people often infer their own attitudes and preferences by observing their own behavior, much as they would infer the attitudes and preferences of others

(Bem, 1967) (Critcher & Gilovich, 2010). In his original experiments, Bem (1967) showed that individuals who were prompted to engage in a certain behavior (e.g., writing essays expressing a contrary attitude) subsequently concluded that they had attitudes consistent with this behavior, even without inner discomfort or cognitive dissonance, which challenged the prevailing theories at the time. Building on this, Critcher and Gilovich (2010) extended the theory to internal mental experiences by showing that people draw conclusions about their attitudes based on the content of their own mind-wandering thoughts. In one of their key studies, participants who thought about environmentally friendly topics during a boring task were more likely to believe they had strong pro-environmental attitudes afterward. In four experiments, they manipulated the content of the mind-wandering thoughts and showed that people interpreted these spontaneous thoughts as meaningful signals of their own beliefs, especially if they did not previously have strong attitudes (Critcher & Gilovich, 2010). Harvey et al. (2007) propose that this might reflect the idea that social or behavioral stimuli engage self-referential processes as individuals try to understand the mental states of others or themselves.

In the context of generating false memories, individuals may incorporate details into their memories based on their beliefs and attitudes, leading to the creation of distorted or inaccurate recollections. According to this, Frenda et al. (2013) found that false memories are implanted and recalled more easily when the false information is congruent with one's attitudes and beliefs, at least in a political context (Frenda et al., 2013; Armaly & Enders, 2023). Additionally, Zuh and colleagues (2010) mention that memories are not accurate representations of past experiences but are influenced by personal meanings and interpretations. The illusory truth effect is another indicator for the assumption that congruent behavior makes false memories more likely, since the truth effect arises because people use peripheral cues such as familiarity

(Begg et al., 1992), fluency, cohesion or congruence as cues for judging truth. Also, Conway and colleagues (2004) show that people prefer coherence and cohesion in memories by showing that people organise autobiographical memories so that they correspond to their current self-image and life story, even if this requires the falsification or omission of details in order to maintain internal consistency and psychological stability. Therefore, we assume in this study that false information in general might be accepted and integrated in one's memory more easily when the context is congruent with the participants beliefs and attitudes in general.

### **Prosocial behavior as a congruent condition**

Contrary to much research that emphasize the selfish or aggressive aspects of human nature (e.g. Zimbardo, 1971; Milgram, 1963), Bregman (2020) and other authors present a vast amount of research and public events that show that humans are inherently prosocial, cooperative and are guided by empathy under typical conditions (Bregman, 2020, pp. 20,21). Building on this view, we hypothesize that human nature is generally pro-social and therefore a pro-social context would provide a congruent situation for participants, leading to a facilitation of false memory creation.

While Richard Dawkins' metaphor of the “selfish gene” popularized the idea that evolution is driven by individual genetic competition (Dawkins, 1976), subsequent biological research and reinterpretations argue for a fundamentally more cooperative framework. As described in Quanta Magazine, genes do not act in isolation but function as an interdependent network that produces the complex mechanisms of life through teamwork rather than conflict. Mark Ridley (2000) argues in *The Cooperative Gene* that genetic success depends on the ability of genes to cooperate, similar to players on a sports team or scientists in a complex project. This interpretation is shared by evolutionary theorists such as Axelrod and Hamilton (1981), who

showed that cooperation can arise and persist even in competitive environments through repeated interactions, and by Nowak, Tarnita, and Wilson (2010), who showed that complex social behaviors can evolve through group and kin selection.

Furthermore, ethologist Frans de Waal (2013) supports the idea that morality, empathy, and altruism are not cultural artifacts, but biologically developed traits with deep evolutionary roots. Although isolated cases of “selfish” genetic behavior, such as transposons or meiotic drive elements, have been documented (Burt & Trivers, 2006), these are the exception rather than the rule. In most cases, genes thrive not by undermining others, but by contributing to the survival and reproductive success of the organism as a whole. This cooperative architecture at the genetic level reinforces the idea that humans are by no means fundamentally selfish but are biologically equipped for prosocial behavior.

Prosocial behavior is not only visible when understanding our genetic evolution, as Kahneman and colleagues (1986) demonstrated that community norms of fairness significantly limit profit-seeking behavior in the markets. Their results show that 82% of the participants found it unfair that a hardware store raises the price of snow shovels from \$15 to \$20 the morning after a large snowstorm, and 77% perceived it as unfair when a company experiencing severe unemployment in the area and decides to reduce current workers' wages by 5%. This preference indicates a concern for equal treatment and social well-being, highlighting an intrinsic motivation to ensure fair outcomes for others and to promote trust and collective well-being by discouraging exploitative practices that harm others. It runs in contrast to conservative economic, which assume that individuals and companies would mainly act out of rational self-interest to maximize their own benefits. To bring the dimension of the negative image of human to a climax, Zimbardo et al. (1971) showed in his Stanford Prison experiment how aggressive, selfish

and immoral people are when feeling powerful in a role and civilization with its rights, values and moral standards is taken away from them. Far less well known is a replication of this study conducted by the BBC, which produced opposite results. Here, researchers did not threaten and manipulate their participants, resulting in pro-social behavior of both prisoners and guards. They did not take their roles very seriously and instead started chatting and playing games with each other (Reicher & Haslam, 2006). To conclude, it appears that pro-social behavior is intrinsic to most human beings and will therefore be considered as a congruent context in the following study.

### **Measuring false memories**

The Deese-Roediger-McDermott paradigm - Very popular in the research field of false memories is the so-called “Deese-Roedinger-McDermott” paradigm (DRM). Here, test subjects are asked to remember words read aloud from word lists on a specific topic. People tend to mentally supplement these lists with suitable, similar words and remember them (cf. DRM effect). For example, test subjects often remember the word “needle” when the word list read aloud covers the topic “doctor” or “vaccination”, even though the critical word “needle” was never presented. However, it is still unclear what exactly is being recorded with this method. Is this really about memories or is the DRM effect due to top-down processes such as schemas, associations and familiarity? Bernstein and colleagues write in their 2018 review: “What do these memory errors have in common? The short answer is, not much. Although researchers have referred to all these errors as false memory, there is little evidence that these errors relate to each other.” (Bernstein et al., 2018, pp. 161,162). In 2011, Otgaar and Candel also found no connection between suggested false events (cf. false event suggestion effect, Loftus & Pickrell, 1995) and false DRM memories (Otgaar & Candel, 2011). The phenomenon of a false memory strongly depends on the

theoretical construct of a memory at hand. These quickly generated “remember and recall” tasks primarily make demands on the working memory. It therefore remains to be clarified whether episodic memory is involved in recall at all and to what extent. Despite all these limitations, it is important to mention that Bamatraf and colleagues 2015 found a significant difference in active brain areas (measured using an EEG) between correct and incorrect responses in the recall test after a DRM task. Here, correct and incorrect answers in the memory test were interpreted as true and false memories, as is usual in the DRM paradigm (Bamatraf et al. 2015). Whether false memories were really recorded or the effect can be attributed to another phenomenon (e.g. familiarity) does not change the fact that the results of the study indicate that there are cognitive differences in retrieval processes.

The post event misinformation effect - The so-called “post event misinformation effect” described by Loftus and colleagues in 1978 is a method with which suggested false memories can be measured more reliably according to our definition of false memories (Loftus et al., 1978). People witness a virtual scene (such as a robbery) and are then asked to remember as many details as possible. In 2005, Okado and Stark developed two “slideshow” of events and had test subjects recall the events using suggestive and non-suggestive questions. In the first event, a purse is stolen from a woman and in the second event, a man breaks into a car (Okado & Stark, 2005). This method has a stronger memory character than the DRM method, as it involves observing an emotional situation and is more likely to create a mental, moving image, which results in a higher probability of integration and retrieval from episodic memory (Kosslyn et al., 1979). In addition, in the second step, the memories are manipulated with critical sentences so that false memories can arise. This method can be used to control the “memory” better, because

we know exactly what was shown in the slideshows and in the misinformation phase and what was not, which is why this study works with the misinformation effect to create false memories.

### **Can we distinguish between true and false memory?**

Even though Loftus et al. states in 1995: “The findings do not, however, give us the ability to reliably distinguish between real and false memories, for without independent corroboration. Such distinctions are generally not possible.” (Loftus & Pickrell, 1995, pp. 755), there are some hints for a possible distinction on a neuroanatomical level: The functioning and neuronal connections of episodic memory have been the subject of much debate for decades. In order to understand whether the episodic memory is activated with a time delay or possibly not at all when a false memory is first generated in the frontal cortex, episodic memory must first be localized. Together with semantic memory, episodic memory forms declarative memory (Renault et al., 2016). The medial temporal cortex (MTC) is most strongly associated with episodic memory in the literature. These deeper areas include the hippocampus, perihippocampal, entorhinal, and parahippocampal cortices, which surround the hippocampus. fMRI studies have shown that the perihippocampal cortex is active during recognition tasks, i.e., when familiarity with test objects is involved (Montaldi et. al., 2006), such as in the DRM-memories. When it comes to memories (not just familiarity), an interaction between the hippocampus, perihippocampal and parahippocampal cortices has been proposed (Montaldi & Mayes, 2010). The hippocampus would reconstruct object information from the perihippocampal cortex and contextual information from the parahippocampal cortex into a memory (Montaldi & Mayes, 2010). According to Pearson's Reverse Hierarchy Model (Pearson 2019), this process should also take place in the case of generating a false memory, but only after the frontal cortex has been active and has generated a mental image of the perception. Parts of the medial temporal

cortex (MTC), as described above, would then compare and supplement this perception with memory content, whereupon the mental image would be drawn in the visual cortex. If an image has already been generated and recalled strongly or frequently, the (false) memory may already have been transferred to the autobiographic memory, meaning that it seems familiar and activates the episodic memory (Pearson, 2019). In this case, the activity of episodic memory, i.e. neuronal evidence of a true memory, could no longer be implemented. However, if, for example, the first interview with an eyewitness is conducted without the eyewitness having been suggestively influenced beforehand and so has already generated specific ideas or images, then, according to the theory, different temporal activations in the frontal cortex and in the MTC should be recognized, as well as differences in the activation of episodic memory, depending on whether the person remembers something true or not (Pearson's theory extended). Even if it seems theoretically possible to find neuroanatomical differences between true and false memories, it remains difficult, if not impossible, to reliably distinguish between them using currently available methods (Loftus & Pickrell, 1995, pp. 755).

## **Conclusion**

Eventhough it remains difficult to disinguish between true and false memories it is evident that false memories exist. The aim of this study is to investigate how the social context of a situation influences the formation of false memories. We assume that participants who are exposed to a prosocial context are more likely to develop false memories. Such a scenario is consistent with widely held beliefs and behaviors, as people generally view prosocial actions as more typical than selfish ones. We expect this effect because previous studies have shown that exposure to congruent information in a political context promotes the formation of false memories (Frenda et al., 2013). To reliably create false memories, we use the misinformation

paradigm. In addition, we include a personality test that measures prosocial traits as an exploratory moderator variable to examine whether individual differences additionally influence susceptibility to false memories.

### References

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. <https://doi.org/10.1126/science.7466396>

Bem, D. J. (1967). *Self-perception: An alternative interpretation of cognitive dissonance phenomena*. *Psychological Review*, *74*(3), 183–200. <https://doi.org/10.1037/h0024835>

Critcher, C. R., & Gilovich, T. (2010). *Inferring attitudes from mindwandering*. *Personality and Social Psychology Bulletin*, *36*(9), 1255-1266. <https://doi.org/10.1177/0146167210375434>

Burt, A., & Trivers, R. (2006). *Genes in conflict: The biology of selfish genetic elements*. Harvard University Press.

Conway, M. A., Singer, J. A., & Tagini, A. (2004). The self and autobiographical memory: Correspondence and coherence. *Social cognition*, *22*(5: Special issue), 491-529.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

de Waal, F. (2013). *The bonobo and the atheist: In search of humanism among the primates*. W. Norton & Company.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, 3(2), 208–214. <https://doi.org/10.3758/BF03212420>

Hyman, I. E., & Billings, F. J. (1998). Individual differences and the creation of false childhood memories. *Memory*, 6(1), 1–20. <https://doi.org/10.1080/741941598>

Loftus, E. F. (1998). The reality of repressed memories. *American Psychologist*, 53(5), 518–537. <https://doi.org/10.1037/0003-066X.53.5.518>

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist*, 48(5), 518–537. <https://doi.org/10.1037/0003-066X.48.5.518>

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361–366. <https://doi.org/10.1101/lm.94705>

- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Murphy, G., & Greene, C. M. (2025). Who Got Lost in the Mall? Challenges in Counting and Classifying False Memories. *Applied Cognitive Psychology*, 39(2), e70044. <https://doi.org/10.1002/acp.70044>
- Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466, 1057–1062. <https://doi.org/10.1038/nature09205>
- Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews. Neuroscience*, 20(10), 624–634. <https://doi.org/10.1038/s41583-019-0202-9>
- Renoult, L., Tanguay, A., Beaudry, M., Tavakoli, P., Rabipour, S., Campbell, K., ... & Davidson, P. S. (2016). Personal semantics: Is it distinct from episodic and semantic memory? An electrophysiological study of memory for autobiographical facts and repeated events in honor of Shlomo Bentin. *Neuropsychologia*, 83, 242-256.
- Ridley, M. (2000). *The cooperative gene: How Mendel's demon explains the evolution of complex beings*. Free Press.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54(3), 182–203. <https://doi.org/10.1037/0003-066X.54.3.182>

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston, MA: Houghton Mifflin.

Szojka, Z. A., Block, S., & La Rooy, D. (2025). Is it time to leave the shopping mall behind? Measurement flaws, plausibility, and external validity of false memory research. *Applied Cognitive Psychology*.

## Appendix 2

### Pretest Results

To analyze pre-coding texts of the pre-test, we conducted 14 independent sample t-tests to compare the ratings of the man's behavior between the social and selfish condition on each adjective of the scale. The mean rating for the item “empathic” (1) in the social condition ( $M = 6.09$ ,  $SD = 0.971$ ) is significantly higher than the mean rating for the same item in the selfish condition ( $M = 2.67$ ,  $SD = 1.572$ ),  $t(38) = 8.45$ ,  $p < .001$ . Just as in the first item we found a significant difference between the social and the selfish text in every single t test of each of the 14 adjectives  $p < .001$ .

An exploratory factor analysis (EFA) with oblimin rotation was conducted separately for the social and selfish conditions to validate the expected two-factor structure of the rating scale. The aim was to assess whether participants' perceptions of the conditions were consistent with the respective constructs - selfishness and sociality. Factors were extracted using the minimum residual method, and the number of factors retained was based on the Kaiser criterion (eigenvalues > 1) and theoretical expectations.

In the social condition, the analysis supported the hypothesized two-factor structure. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.59, indicating moderate suitability for factor analysis, and Bartlett's test of sphericity was significant,  $\chi^2(91) = 283, p < .001$ , confirming sufficient inter-item correlations. Communalities ranged from 0.191 to 0.684.

In contrast, the selfish condition showed a single-factor solution, with a KMO of 0.39, suggesting low sampling adequacy, but Bartlett's test remained significant,  $\chi^2(91) = 163, p < .001$ . Interestingly, the single factor represented a bipolar dimension, with the same items loading strongly positive or strongly negative, effectively aligning the two hypothesized factors on a single continuum. Communalities ranged from 0.138 to 1.00. Despite the statistical one-factor result, the pattern of loadings supports the theoretical distinction between **selfishness** and **sociality**, now situated on opposite ends of a continuum. The results confirmed that the items were appropriately grouped, reflecting the distinct dimensions of the social vs. the selfish construct.

### Appendix 3

#### Main effect of false memory

t-Test für unabhängige Stichproben

		Statistik	df	p		Effektstärke
False memories	Student's t	-0.220	44.0	0.827	Cohens d	-0.0650

Anmerkung.  $H_a: \mu_1 \neq \mu_2$

Deskriptivstatistik für die Gruppen

	Gruppe	N	Mittelwert	Median	Std.-abw.	Std.-fehler
False memories	1	21	3.52	3.00	2.84	0.620
	2	25	3.72	3.00	3.16	0.631

#### Confirmatory Factor Analysis

Faktorladungen

Faktor	Indikator	Schätzung	Std.-fehler	95% Konfidenzintervall		Z	p
				Untere	Obere		
social	Impression_1	1.227	0.173	0.889	1.57	7.11	<.001
	Impression_3	1.014	0.205	0.612	1.42	4.94	<.001
	Impression_10	0.812	0.167	0.485	1.14	4.87	<.001
	Impression_11	0.876	0.167	0.548	1.20	5.24	<.001
	Impression_12	1.111	0.149	0.818	1.40	7.44	<.001
	Impression_13	1.090	0.183	0.731	1.45	5.95	<.001
	Impression_14	1.269	0.162	0.950	1.59	7.81	<.001
selfish	Impression_2	1.336	0.208	0.928	1.74	6.41	<.001
	Impression_4	1.460	0.218	1.033	1.89	6.71	<.001
	Impression_5	0.773	0.307	0.170	1.38	2.51	0.012
	Impression_6	1.642	0.219	1.213	2.07	7.50	<.001
	Impression_7	1.697	0.209	1.287	2.11	8.11	<.001
	Impression_8	1.673	0.219	1.243	2.10	7.63	<.001
	Impression_9	1.552	0.235	1.091	2.01	6.59	<.001

[9]

## Faktorkovarianzen

		Schätzung	Std.-fehler	95% Konfidenzintervall		Z	p
				Untere	Obere		
social	social	1.000 <sup>a</sup>					
	selfish	-0.332	0.142	-0.611	-0.0542	-2.34	0.019
selfish	selfish	1.000 <sup>a</sup>					

<sup>a</sup> fester Parameter

## Test auf exakten Fit

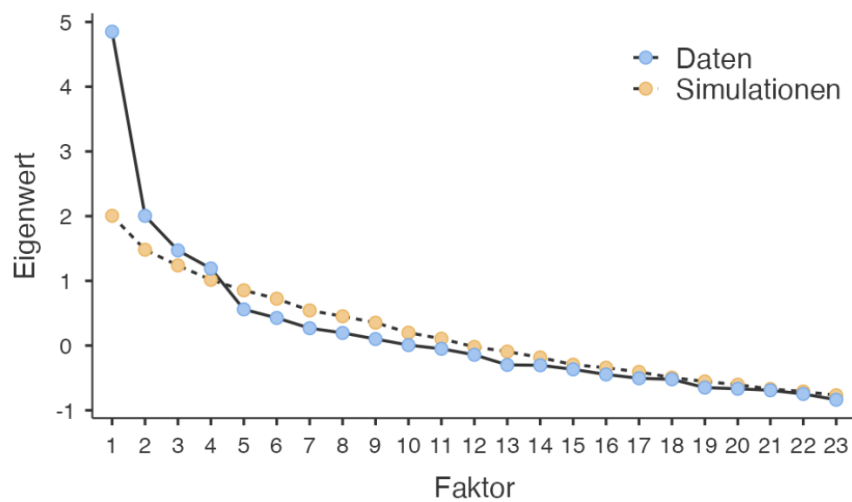
$\chi^2$	df	p
140	76	<.001

## Maße der Modellgüte

CFI	TLI	RMSEA	RMSEA 90% CI	
			Untere	Obere
0.873	0.849	0.137	0.101	0.172

Explorative Factor Analysis

## Scree-Test



## Faktorladungen

	Faktor				1 - Kommunalität
	1	2	3	4	
19 anonymus	0.838				0.291
11 anonymus	0.758				0.392
15 anonymus	0.655				0.509
8 anonymus	0.604				0.631
21 emotional	0.505	0.356			0.438
17 emotioanl	0.390	0.304	0.306		0.496
14 dire	0.380		0.306		0.548
10 altruism	0.316		0.303		0.758
7 compliant		0.824			0.296
18 compliant		0.780			0.369
2 emotional		0.565		0.458	0.447
6Dire		0.550			0.524
9 dire		0.376			0.679
12 emotional	0.364	0.367			0.443
16 altruism			0.765		0.432
23 altruism			0.715		0.477
22 anonymus	0.502		0.543		0.357
13 public					0.907
4 altruism				0.684	0.498
5 public				0.546	0.668
3 Public				0.543	0.538
1 Public	-0.317			0.477	0.713
20 altruism				0.385	0.762

Anmerkung. „Kleinstes Residuum“-Extraktion wurde zusammen mit einer „Oblimin“-Rotation verwendet  
[7]

## Modellanpassung

Güte der Modellanpassung

RMSEA	RMSEA 90% CI		TLI	BIC	Modelltest		
	Untere	Obere			$\chi^2$	df	p
0.0532	0.00	0.0923	0.821	-447	192	167	0.086

## Überprüfung der Voraussetzungen

Bartlett's Test auf Sphärizität

$\chi^2$	df	p
505	253	<.001

## GLMs

### *altruistic*

ANOVA Omnibus tests

	SS	df	F	p	$\eta^2p$
Model	43.769	3	1.7167	0.178	0.109
Altruistic	0.237	1	0.0279	0.868	0.001
condition	0.446	1	0.0525	0.820	0.001
Altruistic * condition	42.906	1	5.0485	0.030	0.107
Residuals	356.949	42			
Total	400.717	45			

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		$\beta$	df	t	p
				Lower	Upper				
(Intercept)	(Intercept)	3.5186	0.434	2.6428	4.394	0.0000	42	8.108	<.001
Altruistic	Altruistic	0.0296	0.177	-0.3283	0.388	0.0245	42	0.167	0.868
condition1	2 - 1	0.1988	0.868	-1.5528	1.950	0.0666	42	0.229	0.820
Altruistic * condition1	Altruistic * 2 - 1	0.7970	0.355	0.0812	1.513	0.6584	42	2.247	0.030

### *dire*

ANOVA Omnibus tests

	SS	df	F	p	$\eta^2p$
Model	39.4569	3	1.5291	0.221	0.098
Dire	0.0978	1	0.0114	0.916	0.000
condition	0.4674	1	0.0543	0.817	0.001
Dire * condition	39.0173	1	4.5361	0.039	0.097
Residuals	361.2605	42			
Total	400.7174	45			

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		$\beta$	df	t	p
				Lower	Upper				
(Intercept)	(Intercept)	3.5227	0.437	2.6417	4.404	0.0000	42	8.069	<.001
Dire	Dire	0.0184	0.173	-0.3305	0.367	0.0157	42	0.107	0.916
condition1	2 - 1	0.2035	0.873	-1.5586	1.966	0.0682	42	0.233	0.817
Dire * condition1	Dire * 2 - 1	0.7366	0.346	0.0386	1.434	0.6284	42	2.130	0.039

*emotional***Model Results**

ANOVA Omnibus tests

	SS	df	F	p	$\eta^2p$
Model	83.0929	3	3.66250	0.020	0.207
Emotioanl	22.8215	1	3.01773	0.090	0.067
condition	0.0218	1	0.00288	0.957	0.000
Emotioanl * condition	39.4003	1	5.20997	0.028	0.110
Residuals	317.6245	42			
Total	400.7174	45			

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		$\beta$	df	t	p
				Lower	Upper				
(Intercept)	(Intercept)	3.5165	0.410	2.6888	4.344	0.0000	42	8.5737	<.001
Emotioanl	Emotioanl	0.1969	0.113	-0.0318	0.426	0.2483	42	1.7372	0.090
condition1	2 - 1	0.0440	0.820	-1.6114	1.699	0.0148	42	0.0537	0.957
Emotioanl * condition1	Emotioanl * 2 - 1	0.5174	0.227	0.0599	0.975	0.6526	42	2.2825	0.028