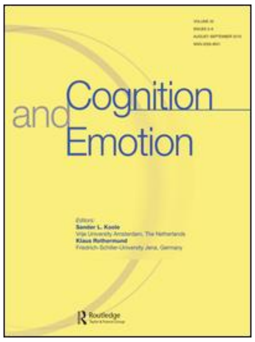

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: https://www.tandfonline.com/loi/pcem20

False memory for pictorial scripted material: the role of distinctiveness and negative emotion

Enrico Toffalini, Chiara Mirandola, Carla De Simone Irace & Gianmarco Altoè

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2020.1749034

BRIEF ARTICLE

False memory for pictorial scripted material: the role of distinctiveness and negative emotion

Enrico Toffalini [a], Chiara Mirandola[a], Carla De Simone Irace[a] and Gianmarco Altoè [b]

[a]Department of General Psychology, University of Padova, Padova, Italy; [b]Department of Developmental and Social Psychology, University of Padova, Padova, Italy

ABSTRACT
Emotional content has complex effects on false memory. Under certain circumstances, emotional material may reduce the likelihood of false memory, a phenomenon that some consider related to it being more distinctive than neutral stimuli. In the present study we tested inferential false memory related to emotionally neutral or negative, and distinctive (but not emotionally charged) scripted material. Remember/familiar judgements were required for recognised stimuli. Data were analysed using mixed-effects multinomial regressions and a Bayesian inferential approach. Results obtained with 82 adult participants showed that, compared with neutral material: distinctive material reduced their false memory associated with "remember" and "familiar" judgements, virtually to the same extent; negatively-charged material reduced false memory associated with "remember" judgements but it had no effect on false memory associated with "familiar" judgements. In short, negatively-charged and distinctive material seems to affect false memory in different ways: the latter affects both recollection and familiarity, the former only recollection.

ARTICLE HISTORY
Received 13 June 2019
Revised 24 January 2020
Accepted 24 March 2020

KEYWORDS
False memory; Emotions; distinctiveness; remember/familiar judgements

Introduction

False memory can arise from reconstructions and inferences based on viewed material, such as scripted events (e.g. Bower et al., 1979; Hannigan & Reinitz, 2001). Recent research has suggested that, when a scripted event depicts emotional scenes, false memory of logically related elements may be reduced by comparison with non-emotional scenes (e.g. Mirandola et al., 2014; Mirandola et al., 2017; Toffalini et al., 2019). It has been suggested that emotion-related false memory reduction is linked to a greater episodic distinctiveness of emotional material, which would facilitate a more detailed encoding. The hypothesis is that a critical element/scene that stands out against similar material around it is protected against memory distortions (Kensinger & Corkin, 2004; Mirandola et al., 2017; Pesta et al., 2001). Unexpectedness may be the key factor, via a "distinctiveness heuristic" (Doss et al., 2019). On the other hand, when series of emotional items are encoded, and critical lures (i.e. non-encoded items that are related with encoded items and may be falsely recalled or recognised during memory retrieval) are also emotional, such a shared "emotional gist" may boost (rather than reduce) false memory by increasing the sense of familiarity (Bookbinder & Brainerd, 2017).

Distinctiveness alone cannot explain what differentiates between emotional and non-emotional (false) memory. Kensinger and Schacter (2008) found enhanced memory for emotional material associated with the activation of a neural network specifically responsible for processing emotional information, that includes the amygdala and the orbito-frontal cortex. Evidence recently reviewed by McGaugh (2018) confirmed that memory consolidation for stimuli encoded before, during or shortly after emotional arousal is regulated by neurobiological mechanisms specifically related to emotional responses, including the release of stress hormones

CONTACT Enrico Toffalini enrico.toffalini@yahoo.it
Supplemental data for this article can be accessed https://doi.org/10.1080/02699931.2020.1749034

and activation of the amygdala. Individual differences in emotional information processing may modulate false memory rates for emotionally charged material as well. For example, individuals with high (subclinical) trait anxiety were more likely than controls to experience false memory for negatively arousing material (e.g. Toffalini et al., 2015; see also Joormann et al., 2009, for similar evidence in the clinically depressed). Such an increase in false memory has been explained in terms of emotional stimuli being elaborated more than neutral stimuli. This suggests, once again, that – beyond the issue of distinctiveness – emotionality per se must also be involved in modulating false memory for emotionally charged material.

In the present study we examined whether distinctive but not emotional material protects against false memory in much the same way as negatively arousing, but similarly distinctive material. An ecological false memory paradigm based on pictorial scripted material originally developed by Hannigan and Reinitz (2001), and recently adapted to include emotionality as a factor (Mirandola et al., 2014, 2017; Toffalini et al., 2019), was extended in the present study to include a new, not emotionally-charged, distinctive condition. In this paradigm, participants encode scripted sequences of photographs depicting everyday life events. Each sequence has alternative, mutually-exclusive endings of different emotionality (e.g. neutral, negative, positive; Toffalini et al., 2019). The "critical lure" in the test is a photograph depicting the unseen antecedent of the encoded ending ("causal antecedent"), which is the same for all possible endings.

In the present study, in addition to a neutral and a negatively-charged condition, we newly added a "distinctive" condition that consisted of salient/unexpected scenes not charged with an emotional valence. An independent group of 28 participants rated all critical scenes (or script endings) on three dimensions: valence, arousal and distinctiveness. We aimed to match the distinctive and negatively-charged emotional conditions in terms of distinctiveness/unexpectedness, but to make them differ in terms of both valence and arousal. For the sake of parsimony, a positive condition was not included (unlike the approach used in previous versions of the paradigm [Toffalini et al., 2019]).

Subjective awareness associated with memory errors was examined to further investigate the nature of false memory associated with distinctive vs negatively-charged events. Participants were required to report Remember/Familiar judgements associated with recognised photographs, where the former term indicated experiences of vivid recall (associated with the retrieval of qualitative details), and the latter a feeling of knowing (Tulving, 1985; Yonelinas, 2002). Both processes can be seen as accessible to consciousness, as participants are explicitly asked to introspect on their subjective awareness, but remembering is associated with a sense of retrieving "qualitative" details about a previous event, while familiarity is more associated with an overall assessment of the "quantitative" strength of the memory (Yonelinas, 2002).

More distinctive stimuli may boost encoding, thereby reducing overall false memory (i.e. reducing the rates of causal errors associated with both "remember" and "familiar" responses). This may apply to both negatively-charged and distinctive material to much the same degree. As mentioned above, however, emotional material undergoes dedicated processing (Kensinger & Schacter, 2008). Mirandola et al. (2014) suggested that negatively arousing material might reduce false memory associated with "remember" responses more than false memory associated with "familiar" responses. Why should this happen? Negatively-charged material undergoes a deeper and longer cognitive processing than non-emotional material, even triggering ruminative processes (e.g. Curci et al., 2013). Since elaboration processes are the basis for inferential errors in the paradigm that we used (Hannigan & Reinitz, 2001; Mirandola et al., 2014), negatively-charged material potentially boosts at least some aspects of false memory. Specifically, it may prevent vivid recollections of unseen causal antecedents (via an enhanced encoding boosted by distinctiveness) but, through an enhanced elaboration of the material viewed, it may increase a false sense of knowing (i.e. being familiar with) associated elements, such as causal antecedents. Given the above-discussed potential role of trait anxiety in moderating the elaboration of negatively-charged material, we also measured this variable in our participants for control purposes.

Method

Participants

A group of 82 undergraduate students ($M_{age} = 22.5$ years, SD = 6.9, 74% females) volunteered to participate in the study. The number of participants is in line with previous studies using a similar version of the same paradigm (e.g. Mirandola et al., 2014, 2017; Toffalini et al., 2019), and in frequentist terms it could be considered appropriate for detecting a small-to-medium effect size using a paired measures design. Setting Cohen's d at .35, α=.05, within-participant correlation $r = .50$, $N = 82$, then power = .88, and exaggeration ratio = 1.08 (Gelman & Carlin, 2014). Written informed consent was obtained from all participants prior to participation. The study followed the ethical standards of the Italian Psychological Association.

Materials and procedure

False memory paradigm

We used a false memory paradigm based on pictorial scripted material (Mirandola et al., 2014, 2017; Mirandola et al., 2020; Mirandola & Toffalini, 2016; Solomon et al., 2019; Toffalini et al., 2014, 2015, 2019). We added a "distinctive" condition alongside the "neutral" and "emotionally negative" conditions of the previously-used versions. The paradigm consists of a series of nine scripted events depicting young actors engaged in everyday life events (e.g. dating a friend, waking up in the morning). Each scripted story is shown in a series of 14 colour photographs on a computer screen. Each of the nine stories has three alternative endings: one is emotionally "neutral" (neutral valence, low arousal), one is emotionally "negative" (i.e. negative valence, high arousal), and one is "distinctive" (neutral valence, low arousal, but with an unexpected element, e.g. a girl wakes up in the morning, opens a cupboard in the kitchen and, instead of food, she find shoes that she quietly begins to polish). Examples of the stimuli are shown in the Appendix. Each participant was shown only one ending for each story. The endings they saw were counterbalanced across participants. Each participant encoded three stories for each type of ending, in a mixed order.

A specific set of critical lures called "causal antecedents" was designed to assess causal errors (Mirandola et al., 2017). They consist of photographs presented at the recognition stage, that depict scenes immediately preceding the critical ending of the script (and they are never shown during the encoding phase). In each script, the causal antecedent is the same photograph for all alternative endings. Another set of lures called "gap-filling distractors" was designed to assess gap-filling errors. They depict scenes that are generally consistent with the script routines, but not actually encoded. The conceptual difference between "causal" and "gap-filling" lures lies in that the former are related to a single, specific event in the script, while the latter are more generally related to the script (Hannigan & Reinitz, 2001). In our case, causal antecedents were specifically related to the pivotal events by means of which the independent variable (labelled "valence" or "emotionality" in previous studies, but here generalised as "type of ending" to include non-emotional distinctiveness) is factorially manipulated. Finally, hits were tested using encoded "target" photographs. Gap-filling distractors and target photographs were designed to be interchangeable, and their roles were counterbalanced across participants. Previous research consistently showed that hits and gap-filling errors were unaffected by the main emotionality/type-of-ending manipulation (e.g. Mirandola et al., 2017; Toffalini et al., 2019). In all, the recognition phase included 9 causal antecedent lures, 27 gap-filling distractor lures, and 36 target photographs.

Trait anxiety

As previously mentioned, trait anxiety was assessed because of its potential relevance to (emotional) false memory (Toffalini et al., 2014, 2015). The trait anxiety scale of the State-Trait Anxiety Inventory (STAI-Y; Spielberger et al., 1996) was used. It consists of a 20-item questionnaire with responses on a 4-point Likert scale (9 items are reverse-scored). Reliability was very good in our sample, Cronbach's α = .91.

Procedure

During encoding, participants sat in front of a computer screen and were told to pay attention to the photographs they would be shown. The nine pictorial scripts were shown one after the other without interruption. The photographs were shown at a rate of 2 sec, with a 2 sec black interstimulus interval.

There was a 15-min retention period after the encoding phase, during which participants completed unrelated filler tasks. Then they engaged in a self-paced recognition phase. They were told they would see another set of photographs that included some they had already seen in the encoding phase and other, new photographs, in a mixed order. Participants were asked to say "yes"/"no" to indicate whether or not they thought they had already seen each photograph during the encoding phase. For each photograph they recognised, they were asked to say they "remembered" it, meaning they had a clear memory of the previous encounter with the photograph, such that they could recollect some qualitative features of that moment (e.g. they could recall something that came to mind when they first saw the photograph). Otherwise, participants were asked to say it was "familiar" if they thought they had seen the same photograph during the encoding phase, but they could not recollect any qualitative features about their encounter with it. After participants had completed the recognition phase, the STAI-Y scale was administered.

Rating phase

An independent group of 28 undergraduate students (M_{age} = 21.2 years, SD = 2.7, 75% females), none of whom subsequently took part in the experiment, served as judges to rate the material on valence, arousal, and distinctiveness. All stories were judged very easy to understand. Valence and arousal were rated using the SAM (Self-Assessment Manikin; Bradley & Lang, 1994), and two separate 9-point rating scales (from 1 = negative valence/low arousal to 9 = positive valence/high arousal). An additional 5-point Likert scale was used to assess distinctiveness, based on the following question: "Please rate how much the ending is salient, i.e. how much it comes as a surprise and captures your attention".

The ratings were analysed using separate mixed-effects linear models for valence, arousal, and distinctiveness, and fitted using the Bayesian method implemented in the "BayesFactor" package of the R software. The ratings were the response variables, the type of ending was the fixed factor of interest, and scripts and participants were random effects. The Bayes Factor (*BF*) was used as a measure of evidence of the fixed effect. Cohen's *d* was reported as the effect size (estimated as the median value calculated from the posterior distributions). Further details can be found in the Supplemental material (Table S1).

There was very strong evidence of the type of ending affecting the valence, arousal and distinctiveness ratings: all *BF*s $>10^{50}$. As concerns valence, the "negative" endings were associated with much more negative ratings than the "neutral" ($d = 1.94$) or "distinctive" ($d = 2.13$) endings, while the latter two received similar ratings ($d = 0.18$). For arousal, the "negative" endings were again associated with far higher ratings than the "neutral" ($d = 1.56$) or "distinctive" ($d = 1.04$) endings, and the "distinctive" endings were moderately more arousing than the "neutral" ones ($d = .52$), possibly due to the unexpectedness associated with distinctiveness being slightly arousing per se. As for distinctiveness, both the "negative" ($d = 1.19$) and the "distinctive" ($d = 1.39$) endings were rated as much more salient/unexpected than the "neutral" endings; the "negative" and "distinctive" endings were similarly rated on distinctiveness, with the latter judged slightly more salient/unexpected than the former ($d = .21$). All these results were in line with our expectations.

Statistical analyses

As the nature of the response variable was discrete and categorical with three options (i.e. "remember"/"familiar"/"no" for the photographs in the recognition phase), multinomial logistic regression was used, i.e. a generalisation of the binomial logistic regression when possible responses fall into more than two categories. See Jaeger (2008) on the advantage of using logistic regressions to model probability when dealing with categorical response variables. Multinomial regression also enabled us to analyze responses within a single statistical model, without fitting different models for alternative response categories. In our case, "no" was set as the reference or baseline category, and the probabilities of "remember" and "familiar" responses were estimated vis-à-vis the "no" responses. For further control, all rates and estimates concerning the "familiar" responses were subsequently corrected using the "independence remember/know" procedure (IRK; Yonelinas & Jacoby, 1995). It consists of dividing the probability of the "familiar" responses by one minus the probability of "remember" responses. This estimates the magnitude of the "familiar" process in the absence of recollection.

The data consisted of a series of measures repeated by participant, so mixed-effects models were used. For the "causal errors" dependent variable, the model included type of ending (within-participant, 3 levels: neutral [baseline], negative, distinctive) as the fixed effect of interest. Previous literature consistently showed that the type of ending (i.e. "valence") affected the probability of causal errors but not gap-filling errors or hits (e.g. Mirandola et al., 2014, 2017; Toffalini et al., 2019). Nonetheless, we tested type of ending as a predictor also for gap-filling errors and hits, for consistency and completeness. The STAI-Y score was also entered as a covariate in all final models to control for trait anxiety. For "causal errors" the interaction between STAI-Y and type of ending was also tested. Random intercepts were set for participants.

A Bayesian analytical approach was used because it offers several advantages, including: the chance to formalise prior knowledge in the analysis; the emphasis on parameter estimation; and flexibility in examining posterior distributions (Kruschke & Liddell, 2018). Prior knowledge was formalised from previous studies that employed the same false memory paradigm (see below). All models were fitted using the "brms" package (Bürkner, 2017) of the R software, which uses the Markov chain Monte Carlo (MCMC) Bayesian estimation method implemented in the STAN programming language. The models were assessed using four chains with 5,000 iterations (and 10,000 iterations for the reported final models). Convergence was assessed with the "Rhat" (potential scale reduction factor on split chains), and was always below 1.01, indicating a good convergence. The Widely Applicable Information Criterion (WAIC; lower is better) was used to assess evidence in favour of the fixed effects of interest by means of model comparisons (Watanabe, 2010). Adopting the criteria suggested by Raftery (1995) for "BIC" (another information criterion with a similar interpretation in a Bayesian context), we considered $|\Delta WAIC| > 2$ as indicating sufficient evidence for a decision in favour of (or against) a given effect, with smaller differences suggesting only "weak" evidence, and $|\Delta WAIC| > 6$ indicating "strong" evidence. When a fixed effect was supported by sufficient evidence, the posterior distributions of its coefficients were examined. The means of the posterior distributions were reported as the point estimates for the model coefficients (*B*s), with 95% Bayesian credible intervals (BCI, calculated with the percentile method) as measures of uncertainty. A BCI excluding zero is interpreted as indicating a probably non-null coefficient. Overlapping was also calculated to compare posterior distributions, using the "overlapping" package (Pastore, 2018) in R.

Definition of prior knowledge

Informed priors were formalised for all effects of interest. The priors were formalised from a pool of observations collected on a total of 248 young adults who served as control groups in four previous studies that employed similar versions of the present false memory paradigm, including at least an emotionally "neutral" vs "negative" condition (Mirandola et al., 2014, 2017; Toffalini et al., 2014, 2015). Mixed-effects logistic regressions were conducted on these data, with experiments and participants entered as random effects, and a set of non-informative priors. The estimated parameters were adapted for the multinomial regression, and served as the informed priors for the present study. The full set of informed priors is available in the Supplemental material (Table S2).

Regarding the effect of the type of ending, none of the previous studies included the "distinctive" condition. Both for theoretical reasons and for the purposes of comparison, the priors describing the effect of the "negative" condition were therefore applied to the "distinctive" condition as well. For the same reasons, identical weakly-informed priors were also used for both "remember" and "familiar" judgements. For the intercepts, we set the priors based on the assumption that the probability of false memory associated with a "remember" judgement would be around half as much as the probability of false memory associated with a "familiar" judgement, in line with descriptive statistics reported in Mirandola et al. (2014). We assumed that the opposite could apply to hits (as they were expected to be subjectively more compelling).

Due to the many assumptions we made, all the above-mentioned priors were set as weakly informative, with $SD = 1.0$. In fact, priors were used mainly as a benchmark, i.e. for the purpose of drawing comparisons. As a sensitivity analysis, we compared the posterior distributions for causal errors with those estimated using a model without any informed priors. We found that the latter had negligible leverage on the posterior distributions (see Supplemental material, Table S2, Figure S1). Uninformed default priors of the "brms" package were used for effects not mentioned above (e.g. for random effects).

Results

Mean proportions of causal errors, gap-filling errors, and hits are reported for descriptive purposes in Table 1. All "familiar" rates are also provided with the IRK correction, which was small to negligible for false memory rates, but rather large for hits.

Causal errors

As expected, strong evidence emerged in favour of a main effect of type of ending on causal errors

Table 1. Mean proportions (and standard deviations) of causal errors, gap-filling errors, and hits, by type of ending (neutral, negative, distinctive) and Remember/Familiar judgements.

Response variable	Neutral	Negative	Distinctive
Causal errors (Familiar + Remember)	.39 (.32)	.32 (.29)	.28 (.30)
Familiar responses	.24 (.26)	.25 (.25)	.18 (.24)
Familiar responses (IRK-corrected)	.30 (.32)	.27 (.29)	.21 (.28)
Remember responses	.15 (.22)	.07 (.15)	.10 (.19)
Gap-filling errors (Familiar + Remember)	.18 (.16)	.17 (.15)	.19 (.17)
Familiar responses	.14 (.14)	.13 (.14)	.14 (.13)
Familiar responses (IRK-corrected)	.15 (.15)	.14 (.14)	.15 (.15)
Remember responses	.04 (.08)	.04 (.08)	.06 (.11)
Hits (Familiar + Remember)	.81 (.16)	.80 (.12)	.81 (.13)
Familiar responses	.26 (.21)	.26 (.21)	.28 (.21)
Familiar responses (IRK-corrected)	.55 (.31)	.53 (.27)	.56 (.25)
Remember responses	.55 (.26)	.54 (.26)	.53 (.24)

Note: For descriptive purposes, responses were averaged by participant before their inclusion in this table, so the standard deviations reflect inter-individual variability in the mean proportions.

(ΔWAIC = −7.80, when it was added to the model that included only the control variable STAI-Y; see details in the online Supplemental material, Table S3).

Figure 1, panel (a), shows the effect of type of ending: causal errors were more likely in the neutral than in the negative or distinctive conditions. "Familiar" judgements were also more likely than "remember" judgements (independently of whether the IRK correction was applied). Different patterns emerged for the "remember" vs "familiar" judgements, however. Causal errors associated with "remember" judgements were just as likely in the distinctive and negative conditions (and less likely in either case than in the neutral condition). Causal errors associated with "familiar" judgements were less likely in the distinctive condition than in either the neutral or the negative condition, however, with the latter two on much the same level. Model coefficients supported the observations emerging from a visual inspection. To be specific, the coefficient for the "remember" judgements, B =−−.98, 95% BCI (−1.59, −.39), was higher than the one for the "familiar" judgements, B = −.11, 95% BCI (−.54, .31), overlap = 6.2%. Conversely, for the distinctive condition, the coefficients for "remember", B = −.67, 95% BCI (−1.25, −.12), and "familiar", B = −.53, 95% BCI (−.98, −.07), judgements largely overlapped, 61.5%. So the difference between the "remember" and "familiar" judgements was greater in the negatively-charged condition than in the other two conditions combined, ΔB = .79, 95% BCI (.16, 1.44). It should be noted that the overall effect of the negative material (i.e. combining the coefficients for the "familiar" and "remember" judgements) was not smaller than the overall effect of distinctive material, B = −.06, 95% BCI (−.48, .37).

Finally, there was moderate evidence of a positive main effect of the STAI-Y scores on the causal errors (ΔWAIC = +2.44, if removed from the final, best-fitting model), B = .24, 95% BCI (.00, .49) (combining virtually identical coefficients for "remember", B = .25, and "familiar", B = .24, judgements). STAI-Y did not

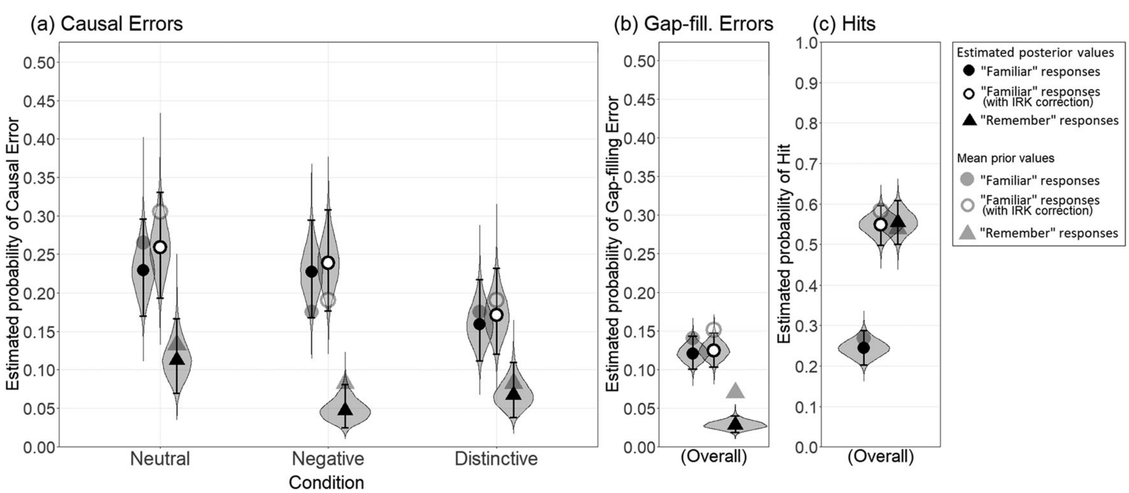

Figure 1. Posterior distributions of the estimated probabilities of causal errors (a), gap-filling errors (b), and hits (c) associated with "remember" vs "familiar" judgements. Causal errors are shown as a function of type of ending. The error bars represent the 95% BCIs of posterior estimates. The grey dots represent the mean values of the informed priors.

interact with type of ending, however (ΔWAIC = +3.04 when the interaction was added to the model).

Gap-filling errors and hits

There was evidence against an effect of type of ending on gap-filling errors (i.e. the fit decreased when it was added to the model including only the control variable STAI-Y, ΔWAIC = +4.05; see details in Supplemental material, Table S3). Likewise, there was evidence against an effect of type of ending on hits (ΔWAIC = +5.47; Supplemental material, Table S3).

The estimated probabilities of gap-filling errors and hits are given in Figure 1, panels (b)(c). As there was substantial evidence against an effect of type of ending on either of these two variables, Figure 1 shows the overall probability estimates. Gap-filling errors associated with "remember" judgements were fewer than those expected from the informed prior, and far fewer than those associated with "familiar" judgements. Hits were largely in line with the priors, more hits being associated with "remember" than with "familiar" responses (the estimated probability of hits, combining "remember" and "familiar", was .80). After the IRK correction, the probability of hits associated with "familiar" responses virtually equalled that associated with "remember" responses (see Figure 1, panel c). Although it was retained as a control variable, STAI-Y was unnecessary as a predictor of either gap-filling errors or hits (ΔWAIC = −0.46 and ΔWAIC = −1.33, respectively, when it was removed from the model; this is not strong evidence against STAI-Y, however).

Additional analysis

An additional analysis was carried out on false memory as a function of the temporal order of lures (gap-filling and causal-antecedent distractors) within the script. This is reported in the Supplemental online material at page 6.

Discussion

Our results suggest that emotionally negative and distinctive (but not emotional) scenes presented in scripts attenuated false memory of causal antecedents in a way that was similar, but not identical. Negative and distinctive material both reduced false memory associated with "remember" judgements, while only the latter did so for false memory associated with "familiar" judgements. We adapted a recent paradigm that allows to investigate a specific, ecologic type of inferential false memory, named "causal errors", i.e. false recognition of the unseen antecedents of seen consequences of varying valences (e.g. Mirandola et al., 2014, 2017; Toffalini et al., 2019).

These findings partly support previous reports of emotional material enhancing encoding, and thus reducing false memory, because it is more distinctive (Kensinger & Corkin, 2004; Mirandola et al., 2017; Pesta et al., 2001). A feature common to such previous research was that the encoded material and critical lures were related, but not by an emotional gist. In the DRM-like paradigm in Kensinger and Corkin (2004), and Pesta et al. (2001), the lures were emotionally charged, but the related encoded stimuli were not. An emotional "distinctiveness heuristic" can explain the tendency to rule out having seen something highly salient/unexpected (such as an emotionally arousing item) on the grounds that it would have been easy to recall (Doss et al., 2019). In our paradigm (as in Mirandola et al., 2017), the "causal antecedent" of the encoded critical event was always inherently neutral. Either way, a mismatch in emotionality/distinctiveness between the encoded and retrieved material seems essential to reducing false memory. In fact, when encoding stimuli share an emotional gist with "critical lures", this may boost false memory (Bookbinder & Brainerd, 2017). The mismatch seems ecologically relevant, however, because crucial facts to remember about an emotionally-charged event might be inherently neutral in real life. This happens in forensics, for instance, when an eyewitness is asked to recall what happened immediately before a crime.

The present study produced additional evidence of how the emotionality of material may affect false memory, after accounting for distinctiveness. We suggest that comparing emotional vs non-emotional, but similarly distinctive conditions is crucial to clarifying the real effects of emotionality. Causal errors associated with our participants' "familiar" responses were no fewer in the emotional than in the neutral condition, suggesting that emotion in itself does not protect against all aspects of inferential false recall. It may be that, although emotional material prompts more specific encoding (e.g. Kensinger & Schacter, 2008), thereby reducing false recall, it does not prevent this material from undergoing further elaboration and deeper engaging processing (Curci et al., 2013), which may trigger schema-based inferential false memory (Bower et al., 1979; Mirandola et al., 2014). Our results are also compatible with the more general notion that emotionally negative material increases verbatim-based encoding without suppressing gist-based false memory (Farris & Toglia, 2019). Distinctive (but not emotionally charged) material was found associated with much the same reduction in false memory associated with "remember" and "familiar" judgements. Another interpretation of our findings is that, although false memories relating to emotionally negative material were clearly no fewer than those relating to distinctive material, they might be subjectively "weaker". This idea could be further investigated in future research by analyzing the confidence level as well as, or as an alternative to remember/familiar judgements (e.g. Kersten & Earles, 2017).

Finally, we had reason to expect a possible interaction between trait anxiety (STAI-Y) and type of ending (i.e. an enhancing effect of the former on causal errors only for negatively-charged content), but our data did not support it. It may be that most participants were in the normal range of anxiety, rather than selected for their higher scores (e.g. Toffalini et al., 2015). Some moderate evidence nonetheless emerged of a positive main effect of STAI-Y (i.e. higher anxiety associated with more causal errors across the board). This seems to contrast with previous reports on false memory from misleading information (Zhu et al., 2010). Trait anxiety may prompt enhanced ruminative processing when encoding stimuli, which is the basis for causal errors in our scripted-material paradigm (Mirandola et al., 2014), but previous reports of an anxiety-related generalised increase in causal errors were inconclusive (Toffalini et al., 2014, 2015).

Limitations of this study include the fact that, given the complex nature of the pictorial scripted material, other factors may be involved, as well as emotionality and distinctiveness. We used an ecological paradigm (Hannigan & Reinitz, 2001; Mirandola et al., 2014) for its potential for shedding light on real-life memory functioning, as discussed above. Future research should establish whether analogue results can also be found with more controlled materials, such as word lists. It will also be important to add a positive condition to fully elucidate the differential effects of emotionality and distinctiveness.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Department of General Psychology, University of Padua, Italy, under research grant-type B, awarded to the first author.

Data availability statement

Data and R code for statistical analysis are freely available on Figshare.com (DOI:10.6084/m9.figshare.8268989)

ORCID

Enrico Toffalini http://orcid.org/0000-0002-1404-5133
Gianmarco Altoè http://orcid.org/0000-0003-1154-9528

References

Bookbinder, S. H., & Brainerd, C. J. (2017). Emotionally negative pictures enhance gist memory. *Emotion*, *17*(1), 102–119. https://doi.org/10.1037/emo0000171

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220. https://doi.org/10.1016/0010-0285

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Curci, A., Lanciano, T., Soleti, E., & Rimé, B. (2013). Negative emotional experiences arouse rumination and affect working memory capacity. *Emotion*, *13*(5), 867–880. https://doi.org/10.1037/a0032492

Doss, M. K., Picart, J. K., & Gallo, D. A. (2019). Creating emotional false recollections: Perceptual recombination and conceptual fluency mechanisms. *Emotion*. https://doi.org/10.1037/emo0000590

Farris, E. A., & Toglia, M. P. (2019). Conjoint recognition procedures reveal verbatim processing enhances memory for emotionally valenced pictorial stimuli. *Emotion*, *19*(3), 533–542. https://doi.org/10.1037/emo0000458

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. https://doi.org/10.1177/1745691614551642

Hannigan, S. L., & Reinitz, M. T. (2001). A demonstration and comparison of two types of inference-based memory errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 931–940. https://doi.org/10.1037/0278-7393.27.4.931

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Joormann, J., Teachman, B. A., & Gotlib, I. H. (2009). Sadder and less accurate? False memory for negative material in depression. *Journal of Abnormal Psychology*, *118*(2), 412–417. https://doi.org/10.1037/a0015621

Kensinger, E. A., & Corkin, S. (2004). The effects of emotional content and aging on false memories. *Cognitive, Affective, and Behavioral Neuroscience*, *4*(1), 1–9. https://doi.org/10.3758/CABN.4.1.1

Kensinger, E. A., & Schacter, D. (2008). Neural processes supporting young and older adults' emotional memories. *Journal of Cognitive Neuroscience*, *20*(7), 1161–1173. https://doi.org/10.1162/jocn.2008.20080

Kersten, A. W., & Earles, J. L. (2017). Feelings of familiarity and false memory for specific associations resulting from mugshot exposure. *Memory & Cognition*, *45*(1), 93–104. https://doi.org/10.3758/s13421-016-0642-7

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–260. https://doi.org/10.3758/s13423-016-1221-4

McGaugh, J. L. (2018). Emotional arousal regulation of memory consolidation. *Current Opinion in Behavioral Sciences*, *19*, 55–60. https://doi.org/10.1016/j.cobeha.2017.10.003

Mirandola, C., & Toffalini, E. (2016). Arousal-but not valence-reduces false memories at retrieval. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0148716

Mirandola, C., Toffalini, E., Ciriello, A., & Cornoldi, C. (2017). Working memory affects false memory production for emotional events. *Cognition and Emotion*, *31*(1), 33–46. https://doi.org/10.1080/02699931.2015.1075379

Mirandola, C., Toffalini, E., Ferruzza, E., & Pazzaglia, F. (2020). The power of extraversion? Reduced false memories for positive events. *Personality and Individual Differences*, *159*, 1–5. https://doi.org/10.1016/j.paid.2020.109861.

Mirandola, C., Toffalini, E., Grassano, M., Cornoldi, C., & Melinder, A. (2014). Inferential false memories of events: Negative consequences protect from distortions when the events are free from further elaboration. *Memory (Hove, England)*, *22*(5), 451–461. https://doi.org/10.1080/09658211.2013.795976

Pastore, M. (2018). Overlapping: A R package for estimating overlapping in empirical distributions. *The Journal of Open Source Software*, *3*(32), 1023. https://doi.org/10.21105/joss.01023

Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 328–338. https://doi.org/10.1037/0278-7393.27.2.328

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Solomon, M., Iosif, A.-M., Krug, M., Nordahl, C., Adler, E., Mirandola, C., & Ghetti, S. (2019). Emotional false memory in autism spectrum disorder: More than spared. *Journal of Abnormal Psychology*, *128*(4), 352–363. https://doi.org/10.1037/abn0000418

Spielberger, C., Pedrabissi, L., & Santinello, M. (1996). *STAI, state-trait anxiety inventory, forma Y: Manuale*. OS Organizzazioni Speciali.

Toffalini, E., Mirandola, C., Altoè, G., & Borella, E. (2019). Inferential false memories for emotional events in older adults. *British Journal of Psychology*, *110*(4), 686–706. https://doi.org/10.1111/bjop.12371

Toffalini, E., Mirandola, C., Coli, T., & Cornoldi, C. (2015). High trait anxiety increases inferential false memories for negative (but not positive) emotional events. *Personality and Individual Differences*, *75*, 201–204. https://doi.org/10.1016/j.paid.2014.11.029

Toffalini, E., Mirandola, C., Drabik, M. J., Melinder, A., & Cornoldi, C. (2014). Emotional negative events do not protect against false memories in young adults with depressive–anxious personality traits. *Personality and Individual Differences*, *66*, 14–18. https://doi.org/10.1016/j.paid.2014.02.042

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*, *34*(5), 622–643. https://doi.org/10.1006/jmla.1995.1028

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., Li, H., Moyzis, R. K., Lessard, J., & 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. https://doi.org/10.1016/j.paid.2010.02.016

Appendix

Note: Example of a script. A boy puts on his jacket, asks his mother for money, and goes to a bar to play slot machines. In the neutral condition, he plays and then goes away. In the emotionally negative condition, he despairs over losing his money. In the distinctive condition, a man comes to and clean the machine screen while the boy is playing. The common causal antecedent (or "critical lure") shows the boy about to insert coins in the machine.