Following the crowd: Brain Substrates of Long-Term Memory Conformity

Micah Edelson[1,*], **Tali Sharot**[2], **Raymond J Dolan**[2], and **Yadin Dudai**[1]

[1]Department of Neurobiology, the Weizmann Institute of Science, Israel

[2]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, UK

Abstract

Human memory is strikingly susceptible to social influences, yet we know little about the underlying mechanisms. We examined how socially induced memory errors are generated in the brain by studying the memory of individuals exposed to recollections of others. Participants exhibited a strong tendency to conform to erroneous recollections of the group, producing both long-lasting and temporary errors, even when their initial memory was strong and accurate. Functional brain imaging revealed that social influence modified the neuronal representation of memory. Specifically, a particular brain signature of enhanced amygdala activity and enhanced amygdala-hippocampus connectivity predicted long-lasting, but not temporary memory alterations. Our findings reveal how social manipulation can alter memory and extend the known functions of the amygdala to encompass socially-mediated memory distortions.

Our memories are often inaccurate. Ubiquitous sources of false recollection are social pressure and interpersonal influence (1-4). This phenomenon, dubbed "memory conformity" (4), is encountered in a variety of contexts, including social interactions, mass media exposure and eyewitness testimony. In such settings an individual may change veridical recollections of past events to match a false account provided by others (1-6). Although these social influences on memory have been extensively demonstrated (1-5), the underlying neurobiology of this process is unknown.

Conformity may present in two forms which initially convey similar explicit behavior but are fundamentally different (7-8). In one type, known as private conformity, an individual's recollection may genuinely be altered by social influence, resulting in long lasting, *persistent* memory errors (1,4-5,7). In such circumstances, even when social influence is removed, the individuals will persist in claiming an erroneous memory as part of their own experience (7,9). Private conformity could hence be considered a *bona fide* memory change. In the second type, known as 'public conformity', individuals may choose to outwardly comply, providing an account that fits that of others, but inwardly maintain certitude in their own original memory. Public conformity can be dispelled when the veracity of the socially transferred information abates (7,10-11). Thus, errors induced by public conformity are *transient* (7,9) and appear to represent a change in behavior in the absence of lasting alterations to a memory engram.

Although private and public memory conformity are often behaviorally indistinguishable, they reflect different cognitive processes (7-8). These processes are probably mediated by

distinct activation in interconnected brain circuits previously found to be active in mnemonic functions and social cognition (such as the hippocampal complex, amygdala and frontal regions) (12-18). Here, we set out to characterize the brain mechanisms that lead to both types of conformity.

Our experimental protocol included four phases spanning a 2 week period (Fig. 1A). Thirty adult participants (12 females, age 28.6 ± 0.8, mean $\pm$ SEM) viewed an eyewitness-style documentary on a large screen in groups of five. Three days after viewing, participants returned to the lab individually and completed a memory test (*Test 1*). *Test 1* served to assess the participants' baseline accuracy and confidence before the manipulation stage. Four days later, participants returned to the lab and answered the same memory questions while being scanned with functional magnetic resonance imaging (fMRI) (*Test 2*). On this occasion a manipulation was introduced in an attempt to induce conformity.

Before responding during this test, participants were presented with answers they were led to believe were given by their four fellow co-observers, whose photographs were provided with their corresponding answers (Fig. 1A). In a subset of trials, for which the target participant originally had a confident veridical memory (as identified by *Test 1*), the answers provided by the four co-observes were all false (*manipulation condition,* 80 questions). In matched control trials the letter X was presented instead of the co-observers answers (*no-manipulation condition,* 25 questions). Pilot data indicated that the use of *manipulation* and *no-manipulation conditions* alone would raise suspicion in the participants' minds that the answers given by the co-observers were fabricated. Therefore we added *credibility trials* in which different patterns of co-observer answers were provided (Fig. 1B).

One week later the participants returned to the lab and were informed that the answers given by the co-observers during the previous fMRI session were in fact determined randomly. This rendered the socially conveyed information previously provided as uninformative. The participants were then requested to complete the memory test again (*Test 3*) based on their original memory of the movie. Finally, the participants were debriefed. Participants with excessive head movements in the scanner or suspected brain pathology and those that indicated suspicion of the manipulation were excluded from the analysis, resulting in a final number of participants (N) = 20.

Our behavioral data revealed that our manipulation induced memory errors (Fig. 2A). Strikingly, participants conformed to the majority opinion in $68.3 \pm 2.9\%$ of manipulation trials – giving a false answer to questions they had previously answered correctly with relatively high confidence. This was not due to forgetting, because in the *no-manipulation condition* incorrect answers were given in only $15.5 \pm 1.7\%$ of the questions ($t (19) = 16.9$, $p < 10^{-7}$). When social influence was removed (*Test 3*), participants reverted to their original, correct, answer in $59.2 \pm 2.3\%$ of the previously conformed trials (*transient errors*) but maintained erroneous answers in 40.8% (*persistent errors*). Confidence ratings in *persistent* and *transient error*s did not differ either before or after the manipulation stage (Fig. 2B). During the manipulation stage, confidence ratings in *transient errors* were significantly lower than in *persistent errors* ($t = 6.9$ $p < 10^{-5}$). Differences in confidence levels were controlled for in the fMRI analysis by means of a covariate (Supporting Online Material, *SOM*).

Our brain imaging data indicated that at the time of exposure to social influence, distinct brain signatures characterized instances of memory conformity that would result in *persistent* and *transient errors*. We first performed analysis on a-priori anatomically defined regions of interest (ROIs) selected by virtue of being widely implicated in memory encoding and maintenance (the bilateral anterior hippocampus, bilateral posterior hippocampus and

bilateral parahippocampal gyrus) and in social-emotional processing (bilateral amygdala) (12-25). Brain activity was averaged across all voxels in each ROI for the three conditions of interest (*persistent errors*, *transient errors* and instances when participants did not conform to the erroneous information, i.e. *non-conformity*). In all regions, except for the left posterior hippocampus, the blood oxygen level dependent (BOLD) signal was greater during trials that subsequently resulted in *persistent memory errors* relative to trials that resulted in *transient errors* or *non-conformity* (Fig. 3A). No significant difference was found between *transient error* and *non-conformity* trials in these regions.

To examine whether other brain regions differentiate between *persistent* and *transient errors,* we conducted a whole brain exploratory analysis. Greater activity during trials resulting in *persistent errors* versus trials resulting in *transient errors* was found in four regions, all in the medial temporal lobe (MTL, Fig. 3B): the left amygdala (−22,−8,−10), right hippocampus (28,−22,−12), right parahippocampal gyrus (PHG, 36,−48,−10) and a region bordering the left PHG and occipital cortex (−22,−54,−10), [$p < 0.001$, cluster threshold (k) > 10]. In the opposite comparison (*transient* vs. *persistent errors*) enhanced activation was found in the bilateral dorsal anterior cingulate cortex (ACC, Brodman Area 32; −12,22,42; 8,20,46).

A striking activation in both aforementioned analyses was found in the left amygdala. A behavioral control study (*SOM*) indicated that elevated activation in the amygdala during trials that resulted in *persistent errors* was not due to heightened emotional arousal during these trials. Nor were these errors related to questions associated with greater emotional content. Rather, heightened amygdala activation seemed specific to socially induced memory change.

The amygdala plays a key role in social and emotional processing and modulates memory related hippocampal activity (13-23). It is strategically placed for this function, having rich anatomical connections with the hippocampal complex (the anterior hippocampus in particular) as well as with neocortical areas (13-16,23,26). The amygdala is thus a prime candidate for mediating social effects on memory, most likely involving its interactions with other brain regions (13-14). This consideration motivated us to carry out a functional connectivity analysis, using a psychophysiological interaction (PPI) approach (27). This analysis showed heightened functional connectivity between the left amygdala and bilateral anterior hippocampus within anatomically defined ROIs, during trials that subsequently resulted in *persistent memory errors* as opposed to *transient errors* and *non-conformity* (Fig. 4A).

We also sought to identify which brain regions responded to the information presented by the co-observers (*SOM*). To this end, trials in which misleading information was presented (the *manipulation condition*) were contrasted with the *no-manipulation condition*. Five regions (Fig. *S1A*) were identified in the frontal and occipital cortex. Further analysis of brain activity in these regions (Fig. S1B) suggests that they are involved in non-mnemonic processes such as conflict monitoring (28-31) in the face of competing memories (32-34).

Were our findings driven merely by the presentation of additional information regardless of social context? To answer this question we performed a control fMRI experiment using a non-social medium to convey misinformation (*SOM*). Participants underwent a similar protocol to that of our main experiment. However, in memory *Test 2*, instead of receiving answers from co-observers, participants were told that the information originated from four different computer algorithms, a common technique used to control for social effects (30). Conformity in this case was significantly lower (45.3 ± 4.7%) than in the *social manipulation* described earlier (68.3 ± 2.9%) but significantly higher than with *no-manipulation* at all (15.0 ± 2.4%) (*t (38)* = 4.2 and *t (19)* = −5.7 respectively; $p < 0.0002$).

Analysis of BOLD signal in the a-priori MTL ROIs revealed an interaction between memory (*persistent errors and transient errors*) and experimental manipulation (*social* and *non-social*) in the bilateral amygdala ($p < 0.05$). This interaction was driven by greater activation in trials resulting in *persistent memory errors* relative to *transient errors* in the *social manipulation*, but not in the *non-social manipulation* (Fig. 3A). These results suggest that enhanced activity in these regions is related specifically to socially-induced *persistent memory errors*. In contrast the right anterior and posterior hippocampus and left PHG revealed a main effect of memory ($p < 0.05$) where there was greater activity during trials resulting in *persistent* relative to *transient errors* regardless of manipulation type ($p < 0.05$) (Fig. 3A). Thus the BOLD signal in these regions was associated with long-lasting memory errors irrespective of the medium by which information was conveyed. Results of a functional connectivity analysis between the left amygdala and bilateral anterior hippocampus showed a significant interaction ($p < 0.05$). Heightened connectivity was seen during trials that resulted in *persistent errors* relative to *transient errors,* a pattern specific to the social manipulation (Fig. 4B). Our control experiment's results hence indicate that heightened amygdala activation and enhanced connectivity with the hippocampus are specific to socially-induced memory changes, whereas hippocampal complex activation differentiates between *persistent* and *transient errors* regardless of the source of influence.

Our results indicate that memory is highly susceptible to alteration due to social influence, creating both *transient* and *persistent errors.* After over a century of intensive behavioral research into social influences on memory (35) this study now provides a brain account of this phenomenon. Our findings suggest a mechanism by which social influence produces long-lasting alterations in memory, and they highlight the critical role of the amygdala in mediating this influence.

Although at the time of social influence on memory overt behavior was indistinguishable, *transient* and *persistent errors* nevertheless induced distinct brain signatures. Heightened activation in the hippocampal complex was seen when false information induced a long-lasting change in the participants' memories regardless of social context. The hippocampal complex activation we observed may represent a process of reconsolidation (36) or encoding of new stable representations (e.g. gist, 37). In contrast transient changes did not activate areas known to be crucial for memory processing. Our findings provide neurobiological evidence for the classic assertion that private conformity is accompanied by actual changes in beliefs whereas public displays of conformity are not (7-8,10,38).

Enhanced activation in the bilateral amygdala and heightened functional connectivity with anterior hippocampus were a signature of long-term memory change induced by the social environment. This indicates that the incorporation of external social information into memory may involve the amygdala's intercedence, in accordance with its special position at the crossroads of social cognition and memory (13-14,16).

Multiple formal models have proposed trace attributes that might contribute to memory distortion in different false memory protocols (37, 39-41). These postulated attributes refer, for example, to potential heterogeneity in episodic content and the persistence of memory trace elements. Our laboratory analogue to socially induced memory distortion was not intended to distinguish between specific models. However, further exploitation of our protocol, combined with cross-fertilization of behavioral and brain data, might contribute to the refinement of current models and better understanding of the biological and cognitive mechanisms of memory conformity.

Altering memory in response to group influence may produce untoward effects. For example, social influence such as false propaganda can deleteriously affect individuals'

memory in political campaigns and commercial advertising (1-2,6) and impede justice by influencing eyewitness testimony (2,4,5). However, memory conformity may also serve an adaptive purpose, because social learning is often more efficient and accurate than individual learning (42). For this reason humans may be predisposed to trust the judgment of the group, even when it stands in opposition to their own original beliefs. Such influences and their long-term effects, the neurobiological basis of which we describe here, may contribute to the extraordinary levels of persistent conformity seen in authoritarian cults and societies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Meade ML, Roediger HL III. Explorations in the social contagion of memory. Mem. Cognition. 2002; 30:995–1009.

2. Loftus EF. Planting misinformation in the human mind: A 30 year investigation of the malleability of memory. Learn. Memory. 2005; 12:361–366.

3. Schacter, DL. The Seven sins of memory: how the mind forgets and remembers. Houghton-Mifflin; New York: 2001.

4. Wright DB, Memon A, Skagerberg EM, Gabbert F. When eyewitnesses talk. Curr. Dir. Psychol. Sci. 2009; 18:174–178.

5. Shaw JS, Garven S, Wood JM. Co-witness information can have immediate effects on eyewitness memory reports. Law Human Behav. 1997; 21:503–523.

6. Perkins HW, Linkenbach JW, Lewis MA, Neighbors C. Effectiveness of social norms media marketing in reducing drinking and driving: A statewide campaign. Addict. Behav. 2010; 35:866–874. [PubMed: 20619177]

7. Smith, E.; Mackie, D. Social psychology. ed. 3. Psychology Press; London: 2007.

8. Allen, V. Advances in Experimental and Social Psychology. Berkowitz, L., editor. Academic Press; New York: 1965. p. 133-170.

9. Reysen MB. The effects of conformity on recognition judgments. Memory. 2005; 13:87–94. [PubMed: 15724910]

10. Asch, SE. Groups, Leadership and Men. Guetzkow, H., editor. Carnegie Press; Pittsburgh: 1951. p. 39-76.

11. Festinger, L. A theory of cognitive dissonance. Peterson; Evanston, IL: 1957. p. 99-100.

12. Dudai, Y. Memory from A to Z. Keywords, Concepts and Beyond. Oxford Univ. Press; Oxford: 2002.

13. Adolphs R. Cognitive neuroscience of human social behavior. Nat. Rev. Neurosci. 2003; 4:165–178. [PubMed: 12612630]

14. Phelps EA. Emotion and cognition: Insights from studies of the human amygdala. Annu. Rev. Psychol. 2006; 57:27–53. [PubMed: 16318588]

15. Dolcos F, LaBar KS, Cabeza R. Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. Neuron. 2004; 42:855–863. [PubMed: 15182723]

16. Dolan RJ. Emotion cognition and behavior. Science. 2002; 298:1191–1194. [PubMed: 12424363]

17. Bickart KC, Wright CI, Dautoff RJ, Dickerson BC, Barrett LF. Amygdala volume and social network size in humans. Nat. Neurosci. 2010; 14:163. (doi:10.1038/nn.2724). [PubMed: 21186358]

18. Ochsner, KN. Social Neuroscience: People thinking about people. Cacioppo, JT., editor. MIT Press; Cambridge, MA: 2005. p. 245-268.

19. Cardinal RN, Parkinson JA, Hall J, Everitt BJ. Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci. Biobehav. R. 2002; 26:321–352.

20. Squire LR. memory systems of the brain: A brief history and current perspective. Neurobiol. Learn. Mem. 2004; 82:171–177. [PubMed: 15464402]

21. Aggleton, JP. The Amygdala: Second Edition. A Functional Analysis. Oxford University Press; Oxford: 2000.

22. Kluver H, Bucy PC. An analysis of certain effects of bi-lateral temporal lobectomy in the rhesus monkey with special reference to 'psychic blindness'. J. Psychol. 1938; 5:33–54.

23. Richardson M, Strange B, Dolan RJ. Encoding of emotional memories depends on amygdala and hippocampus and their interactions. Nat. Neurosci. 2004; 7:278–285. [PubMed: 14758364]

24. Okado Y, Stark CEL. Neural activity during encoding predicts false memories created by misinformation. Learn. Mem. 2005; 12:3–11. [PubMed: 15687227]

25. Nadel L, Moscovitch M. Memory consolidation, retrograde amnesia and the hippocampal complex. Curr. Opin. Neurobiol. 1997; 7:217–227. [PubMed: 9142752]

26. Stefanacci L, Amaral DG. Some observations in cortical inputs to the macaque monkey amygdala: an anterograde tracing study. J. Comp. Neurol. 2002; 451:301–323. [PubMed: 12210126]

27. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage. 1997; 6:218–229. [PubMed: 9344826]

28. Cabeza R, Nyberg RL. Image cognition II. An empirical review of 275 PET and fMRI studies. J. Cogn. Neurosci. 2000; 12:1–47. [PubMed: 10769304]

29. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 2001; 24:167–202. [PubMed: 11283309]

30. Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G. Reinforcement learning signal predicts social conformity. Neuron. 2009; 61:140–151. [PubMed: 19146819]

31. Amodio DM, Frith CD. Meeting of minds: The medial frontal cortex and social cognition. Nat. Rev. Neurosci. 2006; 7:268–277. [PubMed: 16552413]

32. Kuhl BA, Dudukovic NM, Kahn I, Wagner AD. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat. Neurosci. 2007; 10:908–914. [PubMed: 17558403]

33. Levy BJ, Anderson MC. Inhibitory processes and the control of memory retrieval. Trends Cogn. Sci. 2002; 6:299–305. [PubMed: 12110363]

34. Mitchell JP, Dodson CS, Schacter DL. fMRI evidence for the role of recollection in suppressing misattribution errors: The illusory truth effect. J. Cogn. Neurosci. 2005; 17:800–810. [PubMed: 15904546]

35. Bartlett, FC. Remembering. Cambridge University Press; Cambridge: 1932.

36. Dudai Y. The neurobiology of consolidations, or, how stable is the engram. Ann Rev Psychol. 2004; 55:51–86. [PubMed: 14744210]

37. Reyna VF, Brainerd CJ. Fuzzy-trace theory: An interim synthesis. Learn. Individ. Diff. 1995; 7:1–75.

38. Sherif, M. The psychology of social norms. Harper Collins; New York: 1936.

39. Roediger HL 3rd, Watson JM, McDermott KB, Gallo DA. Factors that determine false recall: A multiple regression analysis. Psychon. B. Rev. 2001; 8:385–407.

40. Arndt J. The role of memory activation in creating false memories of encoding context. J. exp. Psychol. Learn. 2010; 36:66–79.

41. Howe ML. What is false memory development the development of? Comment on Brainerd, Reyna, and Ceci. Psycholo. Bull. 2008; 134:768–772.

42. Boyd, R.; Richardson, PJ. social learning: psychological and biological perspectives. Zentall, RR.; Galef, BJ., editors. Erlbaum; Hillsdale, NJ: 1988. p. 29-48.
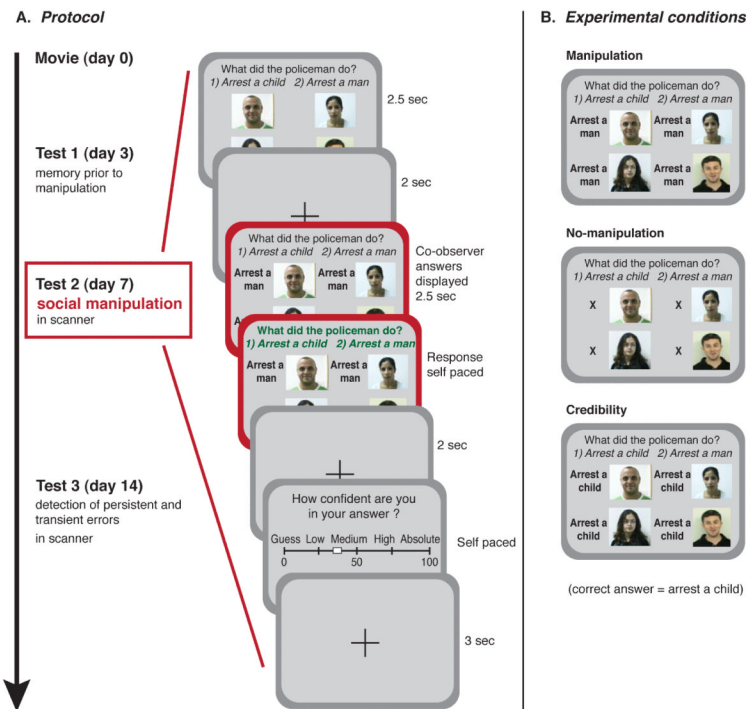
Fig. 1. Experimental Outline

(A) Participants viewed the movie in groups of five, and subsequently preformed three memory tests individually. *Test 1* served to assess the participants' initial memory and confidence before the social manipulation administered in *Test 2*. *Test 3* served to identify memory errors that persisted after the social manipulation was removed. For the *Test 2* scanning session, the question and possible answers were presented for 2.5 seconds, followed by the fabricated co-observers' answers for 2.5 seconds. Subsequently, a font color change indicated that the participants were allowed to respond. Finally, confidence ratings were provided. **(B)** Illustration of the different experimental conditions; the *manipulation condition* in which all co-observers answers were incorrect, the *no-manipulation condition* in which the letter X was displayed instead of co-observers' answers, and the *credibility condition* in which variable patterns of co-observers' answers were displayed (*SOM*).
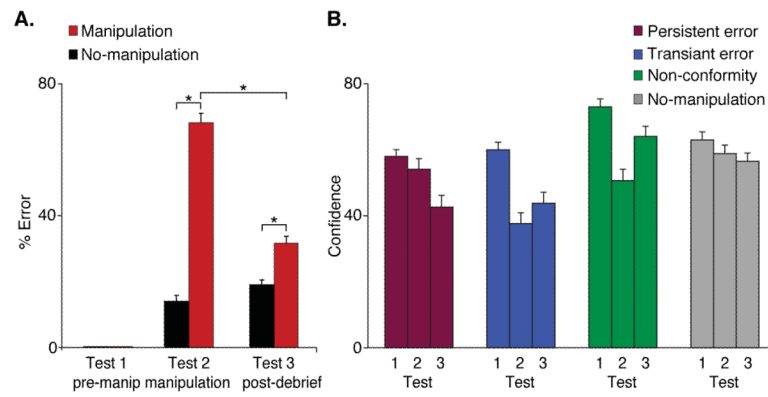
Fig. 2. Behavioral results
(A) Conformity level in the social *manipulation condition* was 68.3% versus 15.5% in the *no-manipulation condition* ($t(19) = 16.9$, $p < 10^{-7}$). In *Test 3* participants reverted back to their original correct answer in 59.2% of the previously conformed-to events (*transient errors*) and on 40.8% maintained their erroneous answer (*persistent memory error*). The error rate was significantly different in *Test 3* between the *manipulation* and *no-manipulation* conditions ($t (19) = 3.7$ $p < 0.002$). The questions included in the *manipulation* and *no-manipulation* trials were those for which participants gave correct answers in *Test 1* with medium-high confidence. **(B)** Confidence ratings over time for differential trial types. (* $p < 0.002$)
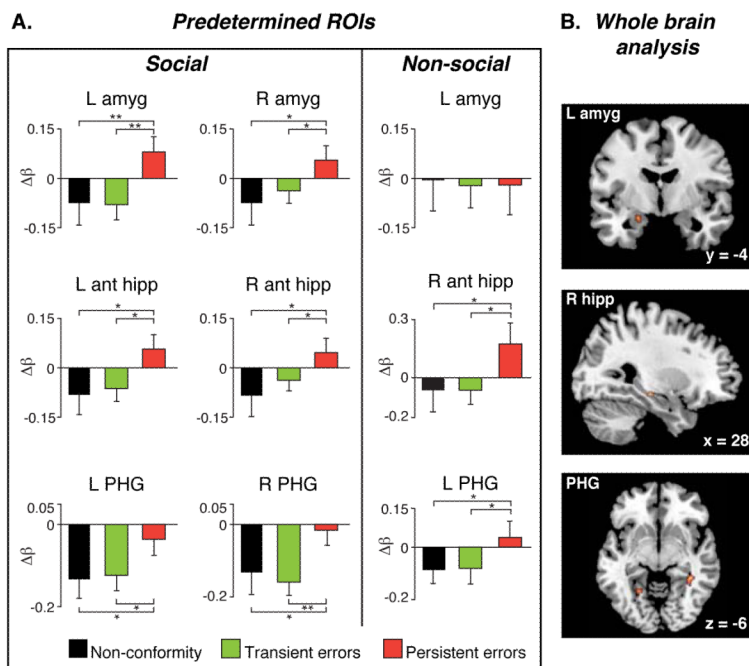
Fig. 3. MTL activation during manipulation predicts long-term socially-induced memory errors
(A) BOLD signal in anatomically a-priori defined MTL regions. L, left; R, right; In the
social manipulation, enhanced activation was found during trials that subsequently resulted
in *persistent errors* relative to all other conditions in the bilateral hippocampal complex and
amygdala. In the *non-social manipulation* this pattern was evident in the hippocampal
complex but not in the bilateral amygdala **(B)** Whole brain exploratory analysis in the *social
manipulation* ($p < 0.001$, k > 10) revealed greater activity in *persistent error* versus *transient
error* in the left amygdala, right hippocampus, right PHG and left PHG bordering on the
occipital lobe. All areas also survived small volume correction for multiple comparisons
(family wise error < 0.05). The baseline in all figures is the *no-manipulation condition*. (* $p
< 0.05$ ** $p < 0.005$)
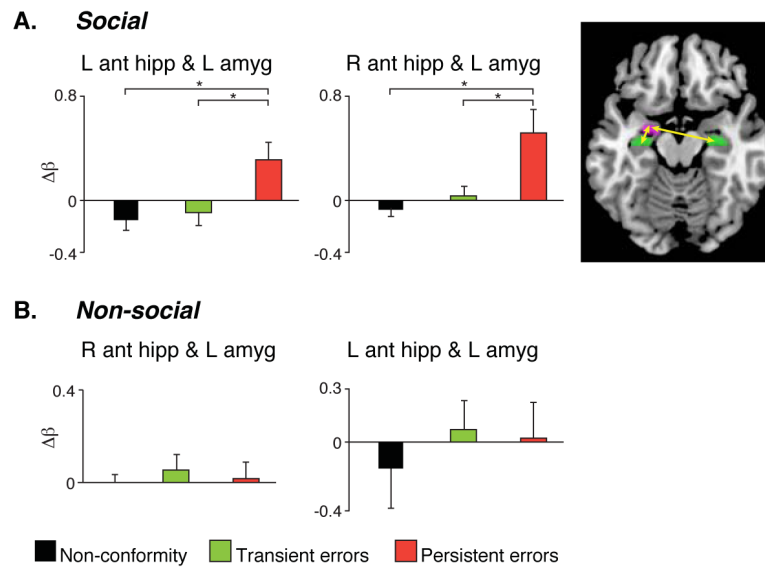
Fig. 4. Amygdala-hippocampal functional connectivity during manipulation predicts long-term socially induced memory errors

(**A**) *Social manipulation.* Functional connectivity between the left amygdala and bilateral anterior hippocampus was heightened in the *persistent error* condition relative to all other conditions. (**B**) *Non-social manipulation.* No condition dependent difference in functional connectivity between the left amygdala and bilateral anterior hippocampus was found. The baseline in all figures is the *no-manipulation condition* (*$p < 0.05$). The inset depicts the anatomical ROIs used in the aforementioned analyses..