Simulating performance in unconscious plagiarism

Nicholas Lange (nick.lange@plymouth.ac.uk)
School of Psychology, Portland Square
Drake Circus, Plymouth University, PL4 8AA
United Kingdom

Timothy J. Hollins (thollins@plymouth.ac.uk)
School of Psychology, Portland Square
Drake Circus, Plymouth University, PL4 8AA
United Kingdom

Abstract

Studies of unconscious plagiarism have reported that people mistakenly include a partner's responses when trying to recall their own (recall-own task) and include own responses when trying to recall their partner's (recall-partner task). In a simulation, we tested if participants' memory performance at test, including source errors, can be explained by participants simply guessing items that come easily to mind. We show that guessing alone cannot account for the pattern of data participants show at test. Modifying the simulation by including memory for self-generated items allows us to replicate the pattern of responding in the recall-own but not the recall-partner task, even when we assume that participants in the recall-partner task strategically withhold more fluent items from report. This suggests that judgements of items' memory strength alone cannot explain performance in the unconscious plagiarism paradigm.

Keywords: source memory; free recall; unconscious plagiarism

Background

In the standard unconscious plagiarism (or cryptomnesia) experiment (Brown & Murphy, 1989), participants in groups take turns to generate solutions for a task. Following a delay participants are asked to complete a recall and/or a generate-new task. In the recall task, participants are asked to selectively recall the solutions they generated themselves, avoiding those generated by others in the group. In the generate-new task, participants are asked to generate novel solutions to the task, avoiding both previously self- and other-generated ones. Plagiarism errors (or source errors in the recall-own and generate-new task) are now solutions generated by other members of the group that participants falsely claim to have generated themselves, with plagiarism typically at above-chance rates for both the recall-own and generate-new task (Brown & Murphy, 1989). More recently, Hollins, Lange, Berry, and Dennis (2016) showed that source errors in recall tasks are not limited to the recall-own task, but also occur during the recall of partner-generated items in the recall-partner task. Rather than participants being biased to simply claim ideas as their own, it appears that participants are simply confused about the source of the ideas they retrieve from memory (Hollins, Lange, Dennis, & Longmore, 2015; Perfect, Field, & Jones, 2009).

While source errors are typically treated as an instance of false memories, an alternative account is that they constitute accidental errors that occur by chance (Brown & Murphy, 1989; Tenpenny, Keriazakos, Lew, & Phelan, 1998). In the study phase, participants are asked to take turns generating responses to cues, such as category exemplars. Without further instruction to generate typical or atypical exemplars, it is likely that participants will first generate responses that are readily available to them, i.e. typical exemplars in the category. This would be in line with participants employing a fluency heuristic (Jacoby, Woloshyn, & Kelley, 1989).

Brown and Murphy (1989) tested this non-memorial guessing account. They presented participants with the test phase of an unconscious plagiarism experiment without a preceding study phase. When treating this generation at test as recall from a study phase that participants did not participate in, participants still committed "source errors" to a high degree. This seems to suggest that reporting items based on fluency or the frequency or typicality of items could be responsible for source errors.

Critically, Brown and Murphy (1989) focused only on source errors, and only on source errors in the recall-own task. In the present paper, we adapted the unconscious plagiarism and anti-plagiarism paradigm as used in Hollins et al. (2015, 2016) for actions. We constructed guessing simulations that builds on Brown and Murphy but attempts to simulate performance across all task measures in both retrieval tasks. If source errors are in part the result of chance performance, the same would have to be true for the correct retrieval and the generation of novel items at test.

We constructed a base guessing model that samples, ignorant to study phase and task, from the possible items per category cue, with sampling weighted by the frequency or typicality of those items. In subsequent simulations, we modified this base guessing simulation by manipulating the number of items per cue available to participants at test, the memory for self-generated and partner-generated items, and the orientation towards self-generated or partner-generated items at test given the retrieval task.

Experimental work

Unconscious plagiarism has been exclusively studied with verbalizable stimuli (for reviews see Perfect & Stark, 2008; Gingerich & Sullivan, 2013). We adapted the unconscious plagiarism paradigm with two retrieval tasks to motor actions to produce observed data. In this experiment, we asked participants to take turns performing and observed actions with a partner in the study phase. In the test phase, participants then were asked to recall performed actions (equivalent to the Recall-own task for verbal material) or asked to recall observed actions (equivalent to the Recall-partner task).

Method

Participants 40 members of the public participated for payment of £12. Three participants did not attend all sessions and their data were excluded from the analysis.

Procedure Participants were asked to attend two sessions in total, a day apart. For the first session, participants were paired and asked to take turns generating and acting out shapes with any part of their body or combinations of body parts. They were shown 15 shape cues (=, A, C, F, H, I, J, K, L, O, P, T, V, X, △) in total. Participants were cued with a printed label of each shape. Members of the pair took turns generating actions for each cue, interleaving performing and observing actions such that performing an action in response to a cue was followed by observing the other person perform an action in response to the same cue. Each participant generated a total of 4 actions per cue, resulting in 60 performed and 60 observed actions overall. Participants were told to observe their partners during partner-generation to avoid duplicating exemplars that had already been created for a cue. Participants observed their partner perform actions under a secondary task load for two-thirds of the shape cues. The assignment of shape cues to secondary task conditions was counterbalanced across participants. The focus of the present paper is on the control condition only, i.e. the one-third of actions participants performed and observed without secondary load, for the purposes of simulating guessing performance.

Participants returned the next day individually for a memory test. They were instructed to retrieve and re-perform either the actions they had generated themselves (Recall performed) or those they had observed their partner perform (Recall observed) the previous day. They were cued with the shape labels, and asked to re-perform as many actions as they could remember for each shape (free report). They were asked to avoid performing actions that did not comply with their retrieval task.

Preprocessing of observed data In the study phase, participants could commit two types of errors: self-plagiarism, that is repeating an item they had already generated for a particular cue, and other-plagiarism, repeating an item their partner had already generated for a cue. In the control condition, participants self-plagiarised on average 5.13% (SD=5.75%) of items and plagiarised 9.73% (SD=6.25%) of partner-generated items. Items that both participants had generated at study for a cue were removed from analysis, since the source of the item, if retrieved, would be ambiguous.

Results

Participants' mean performance in both retrieval tasks is shown by the bars in the figures in this paper. Correct source retrieval was higher in the Recall performed than the Recall observed task, $t(34.80) = 2.75, p = .009$. There was no evidence for source errors or intrusion errors being committed more frequently in one than the other retrieval task, $t(31.86) = 1.29, p = .21$ and $t(35) = .04, p = .97$.

Simulating frequency-based guessing

Base guessing simulation

In the base frequency-based guessing simulation we extended the idea of a test phase without prior study phase tested by (Brown & Murphy, 1989). We used a Monte Carlo procedure to simulate how many correct responses, source errors and intrusion errors participants would make if they were guessing and had just generated potential actions for each shape "on the fly" during the test phase, rather than genuinely retrieving them from what they had previously either seen or performed. We simulated the test phase of the experiment for each participant and each shape separately to take into account differences between individual participants, differing frequency profiles for the different shapes, and the typicality of individual items.

As a first step of the simulation process, we determined frequency norms for the different actions generated for each of the 15 shapes used in the experiment. We took into account all possible shapes all participants had generated across the experiment. Some actions were produced more frequently than others across participants, resulting in a frequency profile for each shape. Both self-plagiarised and partner-plagiarised items at study were included in the creation of these frequency profiles. For each shape, we converted these frequency profiles of the different actions into probability distributions, reflecting the relative probability that a particular action was produced for a given shape. For each shape, the probabilities across the shape cue summed to 1 to represent all possible actions for the shape.

We used these distributions as the basis for participants' guessing. Partner-plagiarised items were excluded from the distribution prior to guessing to match the analysis of the observed data. For each participant we sampled the number of items from each shape distribution that participants reported at test for that particular shape (excluding the partner-plagiarised items again). The sampling was weighted by the relative probability of the items, to implement that guessing was not random but biased by the frequency or typicality and therefore fluency of items. This sampling was done without replacement to match the experimental procedure of only retrieving an item once. Items were sampled sequentially, with the distribution re-normed after each draw. We repeated this for every participant, and ran the simulation over 500 iterations for stable estimates.

In addition to the items participants reported at test (correct responses, source errors, novel items), we now also have the same measures if participants were only guessing at test.

Figure 1 shows the results of the base guessing simulation (the full-length distribution indicated by the stars) relative to the observed data. In both retrieval tasks, mere guessing may approximate the number of source errors but cannot account for the number of correct responses and intrusion errors. It is therefore unlikely that performance in this paradigm occurs simply because highly-frequent items are generated both at study and test, regardless of an influence of memory.

In the next section, we will modify this base guessing simulation by manipulating first the length of the frequency distribution, and then introducing memory and meta-cognition into the frequency-based guessing.

Modifying the base guessing simulation

In the base guessing simulation we assume that participants' responses are based entirely on the overall frequency (or typicality) of items. This assumption leads to the following conditions for guessing: a) each participant has the entirety of each shape distribution available to them at retrieval, b) memory encoding in the study phase does not affect the frequency of items (i.e., there is no effect of memory) and c) participants' responding does not change with the instruction to retrieve items from one or the other source. In the next three steps, we therefore simulated the influences of the length of the distribution, memory effects and retrieval task orientation on participants' guessing. Strictly speaking, only the first modification still represents participants only guessing, i.e., responding without memory. The second modification introduces an effect of memory and the final modification an effect of metacognitive choices made at retrieval.

Length of the distributions In the base guessing distribution, the simulated participants sample their guesses from all possible ways a particular shape was produced in the experiment. This assumes that each participant has access to all possible ways a shape can be represented with the body that were produced throughout the experiment - this is a strong assumption that may inflate the number of novel items relative to items generated at study. It is more likely that each participant has only a subset of items for each shape cue available to draw on. In the first modification of the base guessing simulation, we therefore simulated the pattern of performance in the task if participants guess from frequency distributions for each shape that are a shorter, i.e., include fewer possible actions.

Table 1: Average number of items in the guessing distributions with shortened tails

	Number of items per cue		
Length	Mean (SD)	Max	Min
Full	23.66 (4.68)	33	16
0.9	12.30 (3.19)	19	5
0.8	8.18 (2.20)	13	2
0.7	5.59 (1.26)	8	1
0.6	3.78 (0.88)	6	1

The probabilities associated with the items in the base guessing simulation sum to 1, from most frequent items at higher probabilities to least frequent items with lower probabilities. We created shorter distributions by successively shortening the tail of each shape distribution, i.e., removing the least frequent items. This resulted in distributions representing the top 90%, 80%, 70% and 60% of items generated for each shape across participants in the study phase. The sampling procedure was otherwise identical to the one described above.

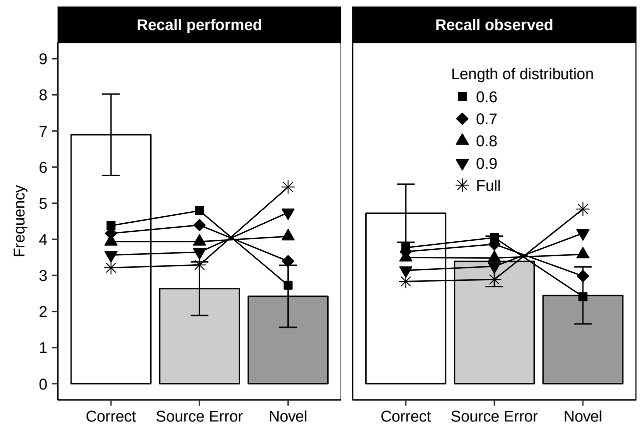

Figure 1: Observed data in bars (correct responses, source errors, intrusion errors) with 95% confidence intervals and data predicted by participants guessing with distributions of varying length relative to the full distribution (points)

Table 1 shows the average number of items, as well as maximum and minimum number of items, that could be sampled across shapes for the different lengths of distributions. For the shorter distributions, in some cases the total number of guesses to be sampled was longer than the distribution to sample from. In those cases, the total possible number of items, i.e. all items in the shortened distribution, was sampled as a guess in lieu of the total number of responses participants in fact made in that case.

Figure 1 shows participants' observed performance (bars) and simulated performance (points and lines) in both the Recall performed and Recall observed task. The stars indicate the sampling based on the full-length distributions for each shape, the remaining points the proportionally shorter distributions (relative to the full-length distribution). Comparing the simulated performance across the different lengths of the distribution shows that with shorter distributions, the number of novel items that are sampled during guessing decreases. There is only a minimal effect on the number of correct responses and source errors that are sampled.

While guessing even based on shorter distributions does not approximate performance in the Recall performed task, guessing based on drastically shortened distributions comes close to replicating the pattern of responding in the Recall observed task. Though note that the radically shortened distributions do not contain many items available for guessing. Naturally, these very short distributions not only contain only a minimum of novel items (hence the decrease in the sampled novel responses), they also do not contain many items participants generated in the study phase and hence do not drastically increase correct and source error responses.

Performance in the Recall-performed task (the original unconscious plagiarism paradigm) cannot be only the result of frequency-based guessing at test. In the Recall observed task, this type of guessing could potentially account for the pattern of responding. In the next step, we modified the distribution further by adding memory for items that were generated at study.

Memory after generation Pure guessing, here implemented by sampling based on the overall frequency or typicality of items, does not fully approximate performance in the memory test and therefore is not an explanation for unconscious plagiarism performance (when both correct responses and intrusions are considered alongside the number of source errors). In the next step, we tested if adding an effect of memory to the model by boosting the probability of items that were generated by participants could account for the pattern of data observed in the experiment.

We used the full-length distribution (rather than shortened distributions). We implemented memory for items by adding a second probability term to all items participants generated themselves (but not to items participants observed their partners perform - we added this modification in the final simulation). The additional probability terms for self-generated items were 0 (the base guessing simulation), 0.1, 0.2, 0.3, 0.4 and 0.5. The final shape distributions were re-normed so all probabilities summed to 1 after this memory probability term was added to the prior probability of each item. Beyond the memory boost, the sampling procedure was identical to the one described in the previous simulations.

Figure 2 shows the observed data (bars) and the predicted responses based on the guessing simulation with memory boost. The memory boost results in good approximation of performance in the Recall performed task (even with the full-length distribution used for guessing). This suggests that participants in the Recall performed task may simply successfully employ a fluency heuristic (Jacoby et al., 1989) by reporting items that are strongly represented at test by a combination of their base typicality and some memory.

In the Recall observed task, increasing the likelihood of generated (here: source error items) to be guessed leads to grave misfits of the pattern of data observed in the experiment. If participants were still simply reporting the most fluent exemplars at test, regardless of the task, the number of source errors (self-generated items with higher memory), should be higher than the number of correct responses (observed actions). This is clearly not what participants in the experiment are doing.

In the final modification, we therefore introduced a metacognitive modification to the simulation that has participants orient their report towards their retrieval task, i.e., deliberately withholding fluent items in the Recall observed task.

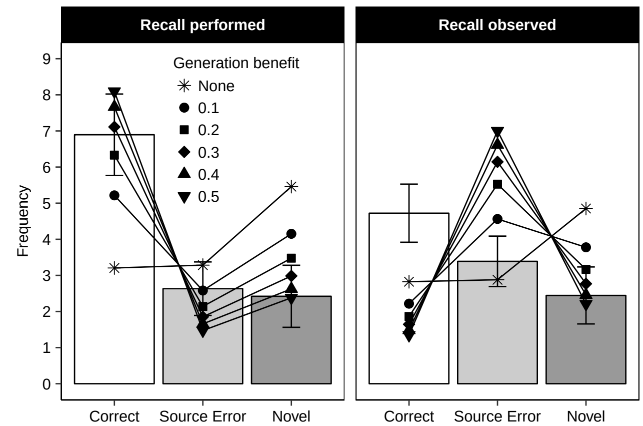

Figure 2: Observed data in bars (correct responses, source errors, intrusion errors) with 95% confidence interval and data predicted by participants guessing with self-generated items' probability to be sampled boosted by varying probabilities.

Orienting towards retrieval task Both the manipulation of the length of the distribution and boosting memory for self-generated items assumed that participants use a fluency heuristic in the test phase of an unconscious plagiarism task. Regardless of the the task instruction to retrieve self-generated or other-generated (here: observed) items, the fluency heuristic assumes that participants will base their responding entirely on what comes to mind at test. This means items associated with higher probabilities will be reported more readily, regardless of the retrieval task.

In a more nuanced approach, it is feasible that participants are able to regulate the memories they report (Marsh & Bower, 1993; Hollins et al., 2016). In this case, participants in the Recall observed task could be able to withhold items that first come to mind from report if they assume that better memory/higher fluency would be indicative of self-generation and hence represents a source error. This type of source monitoring is based entirely on monitoring the memory strength of items, rather than any source features.

In the simulation, we implemented a task orientation by sampling double the total number of items a participant reported from each shape distribution. For the Recall performed task, we then used the top half of the sampled items as items guessed in the simulation. In the Recall observed task, we discarded the first few guesses (this is participants withholding items from report) and instead used the bottom half of guesses, the relatively less frequent items. The remainder of the simulation was identical to previous simulations, with sampling based on the full-length distribution.

Figure 3 shows the results for the frequency-based guessing if self-generated items are boosted in memory and participants in the Recall observed task withhold these items.

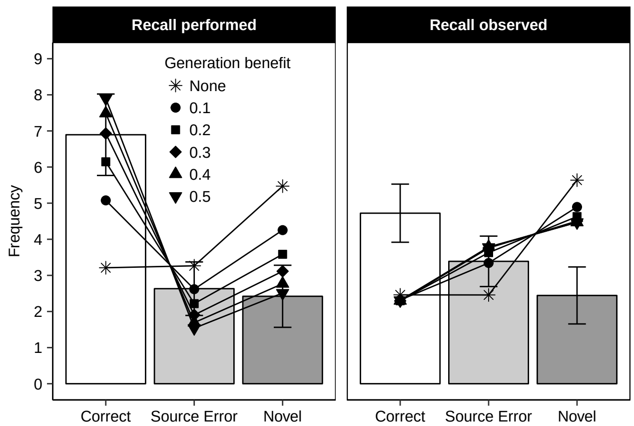

Figure 3: Observed data in bars (correct responses, source errors, intrusion errors) with 95% confidence intervals and data predicted by participants guessing with self-generated items' probability to be sampled boosted by varying probabilities.

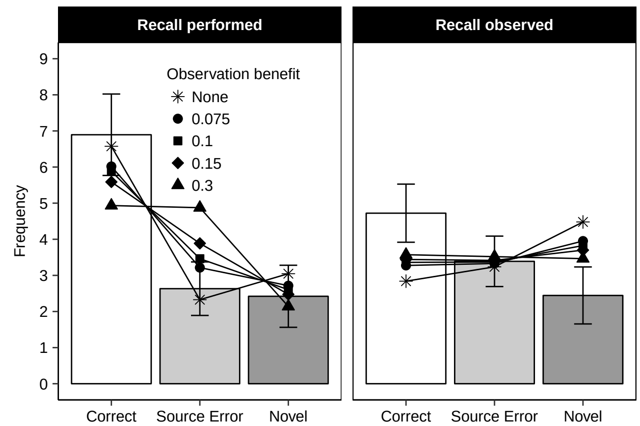

Figure 4: Observed data in bars (correct responses, source errors, intrusion errors) wit 95% confidence intervals and data predicted by participants guessing with self-generated items' probability to be sampled boosted by varying probabilities.

The results for the Recall performed task are naturally equivalent to the simulation without retrieval task orientation, since in both cases the most frequent items are reported. For the Recall observed task, the simulated participants are now less likely to now report source errors, i.e. they successfully withhold those items. Rather than this boosting the correct retrieval (items they observed their partner perform), this modification only increases the number of novel items. Even with an orientation towards weaker items, a frequency-based guessing procedure with memory for self-generated items does not account for the observed pattern of data in the Recall observed task.

Remember, we implemented only increased memory for self-generated but not for observed items. It is possible that even small increases in memory for observed items could explain the correct responses in the Recall-observed task. In a final step modification, we therefore manipulated memory for observed actions. We sampled from distributions slightly limited in length (0.9 distribution from the length modification) and boosted memory for self-generated actions by 0.3 (the boost that most closely matches the pattern of responding in the Recall performed task). We boosted memory for observed actions by a probability term of 0, 0.3/4, 0.3/3, 0.3/2 and 0.3, using the assumption that memory for self-generated items is not likely to be lower than memory for observed actions.

Figure 4 shows the results of the simulation and the observed data. Boosting memory for observed actions does not lead to a closer approximation of the data in the Recall performed task, in part because the effects of the shortened distribution and memory for observed actions both limit retrieval of novel items and increase retrieval of observed actions.

In the Recall observed task, with additional memory for observed actions, the number of correct responses observed in the experiment cannot be replicated. This may not be surprising given the implementation of the task orientation in the simulation. We implemented participants' orientation towards observed actions in the Recall observed task as a withholding of the first few sampled items, in spirit of participants' performance being based on fluency and the interpretation of fluency alone. Increasing the probability of observed items now makes it more likely for those items to be sampled first, and therefore withheld. In other words, to replicate the pattern of observed data using a frequency-based sampling approach, memory for observed actions has to be lower than memory for self-generated actions, with the items participants observed needing to be of higher strength than novel items that were not generated. In the simulations, we did not achieve this balance. It is not clear if memory strength alone is sufficient to explain performance in the Recall observed task.

Discussion

We adapted the extended unconscious plagiarism paradigm (Hollins et al., 2016, 2015) to motor actions. Participants took turns generating and observing actions in the study phase, and were asked to retrieve actions they performed themselves or actions they observed their partner perform in the test phase. We simulated performance in the task to test if guessing alone can account for the pattern of data we observed.

We simulated the experiment to test if frequency-based guessing can account for the observed results. This account is a variation of a fluency or memory strength account of unconscious plagiarism (Marsh & Bower, 1993; Hoffman, 1997; Jacoby et al., 1989) and proposes that memory retrieval in the unconscious plagiarism paradigm is guided by the overall memory strength or availability of items at test. Items with higher memory strength are more likely to come to mind and hence be reported at test. We have shown that this approach with modifications of length, memory and task orientation provides a reasonable description of the data in the Recall performed task. Notably, it does less well in accounting for the data in the Recall observed task.

Given the framework of this kind of memory strength account, the main difference between the retrieval tasks is that participants will do better to report highly-frequent items in one case (Recall performed task) and better to withhold them in another case (Recall observed task). In the Recall observed task, participants ideally report items of some memory strength. In our simulations we were not able to replicate that participants, in fact, are able to make correct responses in the Recall observed task that exceed source errors and intrusion errors.

While responding based on memory strength alone could explain performance in the Recall performed task, it is not sufficient to explain performance in the Recall observed task (for a similar conclusion using a signal detection approach, see Hollins et al., 2016).

There are two possibilities. Participants in the Recall observed task may simply be guessing. In particular if they can only generate very few items (or only very few items beyond self-generated items they remember and are potentially withholding from report), guessing without any memory boost may account for performance in the task. We showed that with very short distributions, performance in the task was approximated. Arguably, the shortest distributions that came closest in matching the pattern afford unrealistically few items to participants in the test phase.

Alternatively, performance in the unconscious plagiarism (and anti-plagiarism paradigm) may not be only based on the overall strength (fluency, familiarity or item memory) of items. In line with the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993) used to explain monitoring failures in other false memory paradigms, participants may judge source memory on a dimension separate to the overall memory strength. While memory strength alone may not allow participants to distinguish very typical items that were not generated from atypical items they observed, retrieving source features from the memory of the observed actions (visual, cognitive, affective, etc.) would allow them to report the observed action over the novel action when asked to do so by the task.

In conclusion, plagiarism errors are not simply the result of participants guessing and reporting typical exemplars at study and at test. While performance when asked to retrieve self-generated items may be explained by participants simply using overall memory strength to guide their responding, performance when asked to retrieve partner-generated items cannot. A source memory account that assumes that participants consider qualitative features of their memory alongside the overall memory strength would be more parsimonious in accounting for performance in both retrieval tasks.

References

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 432–442. doi: 10.1037/0278-7393.15.3.432

Gingerich, A. C., & Sullivan, M. C. (2013). Claiming hidden memories as one's own: A review of inadvertent plagiarism. *Journal of Cognitive Psychology*, *25*, 903–916. (00003) doi: 10.1080/20445911.2013.841674

Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 371–383. doi: 10.1037/0278-7393.23.2.371

Hollins, T. J., Lange, N., Berry, C. J., & Dennis, I. (2016). Giving and stealing ideas in memory: Source errors in recall are influenced by both early-selection and late-correction retrieval processes. *Journal of Memory and Language*, *88*, 87–103. doi: 10.1016/j.jml.2016.01.004

Hollins, T. J., Lange, N., Dennis, I., & Longmore, C. A. (2015). Social influences on unconscious plagiarism and anti-plagiarism. *Memory*, 1–19. doi: 10.1080/09658211.2015.1059857

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of experimental psychology: General*, *118*, 115–125. doi: 10.1037/0096-3445.118.2.115

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi: 10.1037/0033-2909.114.1.3

Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 673–688. doi: 10.1037/0278-7393.19.3.673

Perfect, T. J., Field, I., & Jones, R. (2009). Source credibility and idea improvement have independent effects on unconscious plagiarism errors in recall and generate-new tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 267–274. doi: 10.1037/a0013936

Perfect, T. J., & Stark, L. J. (2008). Tales from the Crypt... omnesia. In J. Dunlosky & Bjork (Eds.), *A handbook of metamemory and memory* (pp. 285–314). New York: NY: LEA.

Tenpenny, P. L., Keriazakos, M. S., Lew, G. S., & Phelan, T. P. (1998). In search of inadvertent plagiarism. *The American Journal of Psychology*, *111*, 529–559. doi: 10.2307/1423550