2016-06

Giving and stealing ideas in memory: Source errors in recall are influenced by both early-selection and late-correction retrieval processes

Hollins, TJ

http://hdl.handle.net/10026.1/5008

10.1016/j.jml.2016.01.004
JOURNAL OF MEMORY AND LANGUAGE
Elsevier BV

Giving and stealing ideas in memory: source errors in recall are influenced by both early-selection and late-correction retrieval processes.

Timothy J. Hollins, Nicholas Lange, Christopher J. Berry & Ian Dennis.

Plymouth University

Date of acceptance: 11.1.2016

Embargo period: 12 months

Note: this is not the document of record: the final version is available at

http://dx.doi.org/10.1016/j.jml.2016.01.004

Acknowledgements

This work was supported by grant funding from the Economic and Social Research Council of the UK (RES-062-23-2766). We thank Chloe Levitt, Daniel Zahra and Amy Richards for their help with data collection, and Chris Longmore and William Simpson for statistical advice. For correspondence about this manuscript, please contact Prof Tim Hollins, School of Psychology, Plymouth University, Drakes Circus, Devon, PL4 8AA, UK.

E-mail: thollins@plymouth.ac.uk

Telephone: +44 1752 584803

Fax: +44 1752 584806

Abstract

Previous studies of unconscious plagiarism have asked participants to recall their own ideas from a previous group-problem solving session, and have typically reported that people mistakenly include a partner’s responses when trying to recall their own. To date, there has been little research looking at the propensity to include one’s own responses when trying to recall a partner’s previous contribution to the group. Experiment 1 demonstrated that people make both kinds of source-error during recall, but source errors are more common in the recall-partner task. This pattern was replicated in Experiments 2a and 2b with source-errors and intrusions increasing over a delay. Experiment 3 used an extended version of each recall task, in which participants reported all items that came to mind, whilst indicating which responses were goal-relevant. The tendency for source-errors to occur more for the recall-partner task was shown to be a function of both idea availability and output monitoring, whereas the tendency for source-errors to increase over a delay was shown to be due solely to output monitoring. Thus, unconscious plagiarism errors are one instantiation of the more general problem of source-specified recall, which is influenced jointly by processes at generation and output monitoring.

Keywords: unconscious plagiarism, source memory, generation; monitoring, recall.

There are many reported disputes about the source of ideas in science, art and everyday life (Defeldre, 2005; Taylor, 1965). These usually feature accusations of intellectual theft accompanied by counterclaims of innocence of intent. Both sides in such disputes could use the experimental literature to support their side of the argument: the accusers can cite the many papers on unconscious plagiarism (also known as cryptomnesia) demonstrating people to be guilty of reporting someone else's ideas as their own. In mitigation, the accused could show that these errors occur even when participants are instructed not to make such errors, or even if they are rewarded for avoiding them (see Perfect & Stark, 2008 for a review). Thus anecdotal accounts and experimental evidence converge to suggest that recall can involve the accidental appropriation of others' ideas, wittily dubbed as *kleptomnesia* by Macrae, Bodenhausen and Calvini (1999). This term implies that unconscious plagiarism constitutes a form of intellectual theft, similar to the self-serving cases of deliberate plagiarism reported in everyday life. We challenge this conclusion here, by showing that people often mistakenly include their own responses when recalling their partner's answers. In fact, giving away ideas in this manner constitutes a higher proportion of recall output than the more commonly studied error of unconscious plagiarism.

Is unconscious plagiarism evidence that we steal ideas from others?

The experimental literature on unconscious plagiarism is dominated by Brown and Murphy's (1989) paradigm. This begins with a group of participants taking turns to individually generate solutions to a presented problem. Later individual group members are instructed to recall their own responses, avoiding those generated by others. Finally, participants attempt to generate new solutions, avoiding all old responses including their own. Brown and Murphy (1989) reported plagiarism at above-chance rates for both the recall-own and generate-new tasks, a pattern subsequently replicated many times (e.g. Brown & Halliday, 1991; Landau & Marsh, 1997; Marsh & Bower, 1993; Marsh & Landau, 1995; Foley, Foley, Durley & Maitner, 2006; Marsh, Landau & Hicks, 1997; Perfect, Defeldre, Elliman & Dehon, 2011; Perfect & Stark, 2012; Stark, Perfect & Newstead, 2005; Stark & Perfect, 2006, 2007, 2008: see Gingerich & Sullivan, 2013 for a recent review). Here we focus only on plagiarism during recall.

The widespread adoption of the Brown and Murphy (1989) paradigm has had the unfortunate consequence of producing an incomplete picture of source-errors in recall. Because the prior work only uses a recall-own task, the only errors detectable involve the mistaken recall of a partner's responses. This methodology doesn't measure whether a participant fails to output one of their own responses because they believe it came from their partner. Nor does it measure the extent to which people would give away their own responses when attempting to recall their partner's responses. Consequently, the literature is replete with examples of individuals claiming other's ideas, but contains almost no examples of ideas being given away or withheld.

This focus on people's tendency to plagiarise the work of others is consistent with social research demonstrating egocentric bias in recall of work with partners. For example, Ross and Sicoly (1979) ran a series of studies in which individuals judged responsibility for group work, such as which spouse contributes more housework (Experiment 1) or how much input a supervisor and a student have on a thesis (Experiment 5). They reported a general egocentric bias, such that people attributed the majority of joint work to themselves, such that the combined input from both partners often exceeded the total amount of work. (For similar examples of egocentric bias in recall see Stephenson and Wicklund, 1983; Hyman, Roundhill, Werner & Rabiroff, 2014). The notion of an egocentric bias in recall of input is also consistent with the many cases of plagiarism of literary or musical content reported in the media. In contrast, to our knowledge there are no reported cases of artists or writers accidentally (rather than deliberately) claiming their own work being by someone else. In this context, it is perhaps understandable why some have claimed that unconscious plagiarism errors are an example of an egocentric bias (Macrae, Bodenhausen & Calvini, 1999; Wicklund, Reuter & Schiffmann, 1988). However, this conclusion is not supported by research on bias in source judgements.

Is there a bias towards plagiarism in source-monitoring judgements?

Memory bias toward or away from the self has been studied most extensively in the source monitoring literature. A common approach in such studies is to have participants presented with one set of items, whilst generating or imagining another set (e.g. Hashtroudi, Johnson & Chrosniak, 1989). Later at test, participants are presented with old and new items, and asked to judge each item as to whether it is old or new, and to judge its source. These two judgements are made either independently in two phases (e.g. Starns, Hicks, Brown & Martin, 2008), or in a single combined decision (e.g. Hashtroudi et al, 1989). Three findings are commonly reported. First, memory is superior for items generated rather than perceived, consistent with the generation effect. Second, people confuse perceived items as having been generated and generated items as having been perceived. Third, falsely recognised new items are most often attributed to the external source – the *it-had-to-be-you* effect (Johnson & Raye, 1981; Hashtroudi et al, 1989; Hoffman, 1997).

Marsh, Landau and Hicks (1997) demonstrated all these effects in a series of experiments looking at the rate of source-errors in recognition tasks involving new items, items generated by the participants, and items generated by other group members. In all experiments, participants were more likely to mistakenly label their own responses as coming from their partners than to label a partner's responses as their own. That is, old items were subject to the same *it-had-to-be you* bias as new ones. However, a complicating factor in these experiments was that the participants originally generated the answers in groups of varying sizes, and so any individual only generated a minority of items. Thus, the tendency to attribute items to another group member could result from a heuristic reflecting the base-rate probabilities. In the absence of source-specifying information, an item is most likely to have come from someone else. Consistent with this observation, the relative rates of plagiarism (claiming someone else's response was self-generated) and anti-plagiarism (attributing one's own response to someone else) across experiments reflected the relative size of the groups involved. Those experiments in which participants generated a higher proportion of the original ideas resulted in higher rates of claiming other's ideas and lower rates of giving ideas away. Nevertheless, in all studies, giving away ideas was more frequently observed.

The one study to date to look at these two forms of source error in a recall-based paradigm was by Perfect, Field and Jones (2009). Participants initially worked with a partner to generate solutions to problems concerning health and the environment, and so base-rate generation was controlled. At the start of the experiment, the participant's partner (a confederate) declared themselves to be an expert in one of the two topics under discussion. After the generation phase, participants thought of ways of improving half the ideas from each topic, before later recalling either their own ideas or their partner's ideas, and then finally generating new solutions to the same topics. With respect to the recall tasks, the results demonstrated more correct recall for ideas generated than for ideas heard from a partner, consistent with the generation effect (Slamecka & Graf, 1978), regardless of expertise (see Perfect, Field & Jones, 2009, Table 1). There was no difference in the frequency of source errors across recall tasks overall, but this was qualified by an interaction with improvement; for the unimproved ideas, giving away ideas (in the recall-partner task) was more frequent than plagiarism (in the recall-own task). For improved ideas, the pattern was reversed. Thus, this study suggests that source-errors occur in both recall tasks, but is equivocal about which form of error is most likely to occur. However, it is unsafe to draw firm conclusions on this basis, because the study involved a manipulation of expertise and an elaboration phase during which half the ideas were improved.

To summarise, the recall-based unconscious plagiarism literature has almost exclusively focused on one form of error, namely the tendency to steal ideas from others. In contrast, the recognition-based source-monitoring literature has measured both intellectual theft and generosity, and suggests that giving away ideas is more common than stealing them. Only one recall-based study has looked at both forms of source error, but the results were inconclusive. Consequently a major focus of the present work is to use recall-based methods to measure the tendency to plagiarise a partner or give away our own ideas to them. A clear prediction from the source-monitoring research is that giving away ideas should be more common than stealing them. However, this prediction neglects key differences between recall and recognition.

Methodological differences between recall and recognition make strong predictions hard to sustain. In recognition tests, the experimenter provides the participants with the set of items to be judged, whilst in recall, the items are output by the participants. Thus, in recall, source-judgements are made only for a subset of items which may not be representative of the full-set of ideas originally generated. In free-recall, participants are likely to recall the strongest memories, failing to access the weakest, and choosing not to report intermediate-strength memories that are accessed with little confidence (Koriat &

Goldsmith, 1996). In contrast, recognition tests require source judgements about all the items (or a randomly selected subset of them) and so include many which would be unlikely to be output during recall. Given that we know that people falsely attribute new items to an external source, it is perhaps not surprising that they make the same errors for old items that they remember very poorly. But this does not make it certain that they would make the same attribution for responses strong enough to be recalled.

A second complication flows from the same issue. In recognition-based source-memory tests, source-errors are considered as a function of the original source, that is, as an input-bound measure of performance (e.g. Marsh, Landau & Hicks, 1997). However, unconscious plagiarism research based upon recall has typically looked at the proportion of output on a given task that comes from the wrong source (e.g. Stark, Perfect & Newstead, 1995; Tenpenny, Kelriazakos, Lew, & Phelan, 1998). As Koriat and Goldsmith (1994) discussed, these different approaches to methodology and analysis can have profound impacts upon the theoretical conclusions drawn.

Early selection and late-correction processes in recall

In addition to the methodological issues, there is a more fundamental distinction between recall and recognition based tests. As discussed above, in a recognition-based test, each item is generated by the experimenter. Consequently, differential rates of source-errors cannot be attributed to any generation process. In contrast, in a recall-based test, the participant must first generate each item before engaging in any source-monitoring judgement of it. That is, the likelihood of a source error in recall is driven by retrieval processes as well as by source-monitoring, but source errors in recognition are only driven by the latter.

Recently there has been growing interest in the interactive roles of generation and monitoring as means to achieve quality control in memory. For instance, consider the examples of trying to recall what you wore to the last party you attended, versus recalling what you wore to the last *fancy-dress* party you attended. The former case is a less specific retrieval cue, and so you may have generated several potential responses, and had to carefully monitor them to evaluate which option was most likely to be correct. Jacoby, Kelley and McElree (1999) call this form of control late-correction. In contrast, for the fancy-dress example, it is more likely that very few responses came to mind, and so very little evaluation was required. In Jacoby, Kelley and McElree's (1999) terminology, this is control by early selection. The role of monitoring in control of memory quality may vary as a function of the specificity of the retrieval process that can be run. One way to enable a specific retrieval process is to engage in distinctive encoding (Halamish, Goldsmith & Jacoby, 2012). Recently, Hunt and Smith (2014) looked at the role of generation and monitoring processes in the ability to selectively recognise items from one of two studied lists, which were encoded either using the same encoding instructions (category membership or pleasantness ratings for each item) or different encoding instructions for each list. In one experiment, participants made recognition judgements for items from both lists either under self-paced or speeded conditions. Speeded recognition removed the ability to distinguish the two lists if they had been encoded using the same instructions, suggesting that late monitoring was required to solve the task, but insufficient time had been given to achieve it. However, participants were still able to distinguish the lists that had been encoded with different instructions under speeded testing. Hunt and Smith (2014) concluded that the distinctive encoding enabled participants to constrain their retrieval to the target list, thus avoiding false positive errors for the non-target list, despite the lack of late monitoring.

The relevance of this early-selection late-correction framework to the task of recalling items from one source or the other is clear. The instruction to freely recall their own or their partner's ideas may result in differences in the specificity of the retrieval process, or the engagement of monitoring. There are good a priori reasons to suspect that both may be the case. For example, the benefits of internal generation over external perception have been attributed to increased distinctiveness (Kinoshita, 1989), and so one might expect self-generated items to be more distinctively encoded in memory. Thus, recall of own ideas may be more amenable to a narrow early-selection process. Additionally, responses from each source may also be subject to different levels of output monitoring. According to the Source Monitoring Framework output monitoring processes are effortful, and may not be adequately engaged if the primary retrieval task is too demanding (Johnson, Hashtroudi & Lindsay, 1993). Given that it is harder to recall a partner's ideas than one's own, a clear prediction is that less monitoring will occur during the recall-partner task.

Consequently, we predict that participants will be more likely to include their own ideas when recalling their partner's ideas than to include their partner's ideas when recalling their own ideas. However, to date, no one has tested the idea in a simple paradigm, and so in the present work we return to the methods used in the original demonstrations of unconscious plagiarism by Brown and Murphy (1989), but include a recall-partner condition to test the tendency to make source errors of donation of own ideas. To avoid any potential influence of base-rate differences in the likelihood of responses being attributed to self or others, we used dyads taking turns to generate items rather than groups. This enabled us to determine the relative strength of any bias towards or away from the self in the absence of any heuristic bias towards the most likely source. Pairs (or virtual pairs) have often been successfully used in research on unconscious plagiarism (Marsh & Bower, 1993; Landau & Marsh, 1997; Macrae, Bodenhausen & Calvini, 1999; Perfect, Field & Jones, 2009).

Four experiments are reported: Experiment 1 provides a straightforward comparison of source-errors in recall of own and partner responses. Experiment 2a extends this work by introducing a delay prior to recall and follow-up Experiment 2b extends by having participants recall both sources simultaneously. Experiment 3 extends the paradigm further by directly exploring the role of early retrieval processes and late monitoring processes in the rate of errors in the two retrieval tasks.

Before we present the experiments, we need to explain our methodological and analytic approach. Our interest in errors made during free-recall requires that participants are free to report what they wish in response to the cues and instructions provided. This distinguishes our approach from recognition-based measures of source-memory as discussed above, but it also distinguishes it from some earlier work on unconscious plagiarism that used forced report recall (e.g. Brown and Murphy, 1989; Marsh & Bower, 1993). That is, if a participant had originally generated four responses to a cue, they were required to recall four at test. As Perfect and Stark (2008) pointed out, this results in many low confidence responses being generated, and questions whether such responses truly represent wrong-source errors. Participants may seek to comply with forced-report instructions by giving responses they know to be source errors.

Consequently, in all but one of the studies reported here, participants were asked to retrieve responses from one source only, accompanied by a warning to try to avoid answer from the alternate source. However, beyond this instruction, they were free to report what they wished. Whilst this does not preclude the possibility of errors made knowingly, it does reduce their likelihood. However, the use of this method complicates the analysis and interpretation of source errors, because both memory strength and report criterion are likely to vary across the recall-own and recall-partner tasks. Figure 1 illustrates the different measures generated by this methodology. In line with previous research (e.g. Perfect, Stark & Newstead, 2005; Stark & Perfect, 2007, 2008; Perfect, Field & Jones, 2009), we report both the absolute rate of unconscious plagiarism errors and the proportion of a participant's output that these errors represent. This latter measure is the one reported by Brown and Murphy (1989) in their original demonstration, albeit they used a forced-report criterion that was fixed for all participants, such that proportion of source errors is equivalent to absolute rate. Note, we prefer to include intrusion errors in our report of source error rate, because we are interested in the rate of errors in a participant's output, rather than the rate of old ideas that wrongly intrude (excluding new ideas). However, if we had adopted this alternate measure of source error rate, all the conclusions reached regarding the effects of task (and delay) on source errors in the Experiments reported here would remain unaltered.

The analyses of source errors described in the previous paragraph are contingent on task output: that is they look at output errors given the retrieval task undertaken, in line with all previous work on unconscious plagiarism. However, our novel use of recall-partner task provides the opportunity to look at wrong-source errors contingent upon original source, in line with recognition-based source monitoring studies. That is, we might wish to ask what proportion of ideas from one source end up being wrongly recalled (during recall from the alternate source), compared to being correctly recalled (during recall from the correct source). This approach is closer to that used in source-recognition work, in that it looks at errors contingent upon original source, but it deviates from that procedure in that it only looks at items recalled, not all the original ideas. However, to formally analyse such data requires that participants recall ideas from both sources during an experiment, which is contrary to work in the unconscious plagiarism literature. In the present set of studies, participants recalled from both sources in Experiments 1 and 2b, and so we report the source-contingent analyses in those experiments. In the other Experiments (2a and 3) we can only discuss the pattern in the group means.

We begin with a straightforward comparison of performance on recall-own and recall-partner tasks. Our prediction, based upon the source-monitoring literature, was that participants would attribute weakly-recalled ideas to a partner, and so be more likely to output errors during a recall-partner task, than during a recall-own task, in line with the performance seen in the control condition of Perfect, Field and Jones (2009)

Experiment 1

Method

Participants: Forty undergraduate volunteers from the University of Plymouth participated on a voluntary basis or for partial course credit. In this, and all subsequent experiments, participants signed up to a time-slot on a web-based participation system, and experimental dyads were created by pairing the two individuals who signed up to the same time slot.

Procedure: Participants, working in pairs at their own rate, alternated in orally generating responses to 2-letter word-stem cues (*BE, TH, FO, MA, ST, RE, SP, PE*), until each had generated 6 exemplars to each of 8 cues. The experimenter verbally introduced each cue in turn, and prompted each participant to respond alternately, writing down their responses as they spoke. Exact repetitions, or derivations of previously presented words (e.g. Fools or Foolish following Fool) were corrected and replacement responses sought, though these errors were not recorded and so frequency data for these corrections are not available. Once all legitimate responses had been generated for a given cue, the experimenter introduced the next cue. The order of generation alternated across cues, such that each participant spoke first for half of the cues. A 3-minute filled interval followed in which participants individually completed Sudoku puzzles. At test, participants were given 2 recall sheets each divided into four quadrants. Each quadrant contained an orthographic cue, and prompted up to 6 responses by means of 6 asterisk symbols. Participants were instructed to freely recall their own answers for one response sheet (containing one set of four cues), and their partner's answers for the other response sheet (containing the other set of four cues). In the recall-own task, participants were verbally told to recall only the answers that they had previously generated, and to try to avoid answers that their partner had previously generated (with the source instructions reversed for the recall-partner task). The assignment of cues to recall task, and order of recall tasks was counterbalanced across participants. Thus, the maximum number of correct responses on each recall task was 24, with the potential to recall the same number from the wrong source, with the same for novel intrusions, with the caveat that the total number of responses overall was capped at 24. For example, to the cue *BE___,* a participant might recall *beat, bell, bean, bed, and bend*. The experimenter then coded these responses by original source to determine the rate of correct recall (e.g. *beat, bell, bed*), wrong-source errors (e.g. *bean*) and novel intrusions (e.g. *bend*). These three response types were then summed across cues, as appropriate. Participants were given 4 minutes to complete the recall task.

Results and Discussion

Table 1 shows how performance varied as a function of the recall-task. Responses were scored as correct if they duplicated a previous verbal response from the correct source, regardless of spelling or plurality (e.g. if a partner had said "bean" to the cue BE____, we would have counted all the following responses as correct: been, bean, beans). We used the same criteria to score wrong-source errors. All responses that could not be traced back to one of the two sources were classified as intrusion errors. We do not report means by order of tasks, because it had no impact upon performance, but we included this factor in our analyses.

Correct recall was higher for the recall-own task, $F(1,38) = 75.82$, $p < .001$, MSe = 8.13, partial eta^2 = .666, but was not influenced by the order of the recall tasks, $F(1,38)$ = 1.62, $p = .21$, MSe = 12.37, partial eta^2 = .041, and there was no task by order interaction, $F < 1$. Wrong-source recall was higher for the recall-partner task, $F(1,38) = 5.46$, $p = .025$, MSe = 1.67, partial eta^2 = .13, but was not influenced by task order, $F(1,38) = 2.38$, $p = .13$, MSe = 5.05, partial eta^2 = .059, and there was no task by order interaction, $F < 1$. For novel intrusions, there was no reliable tendency to attribute to one source over another, $F(1,38)$ = 2.90, $p = .10$, MSe = 3.63, partial eta^2 = .071, and nor was there an effect of task order $F < 1$, and no task by order interaction, $F < 1$.

The analyses above were based upon the absolute rate of wrong-source errors, but this neglects the context of lower overall recall in the recall-partner task. Consequently we also calculated the output-bound proportion of wrong-source errors, that is, the proportion of all items generated during retrieval that came from the wrong source. Analysis of the wrong-source error rates calculated in this manner confirmed the greater propensity to give away answers than to steal them, $F(1,39) = 18.9$, $p < .001$, MSe = .012, partial eta^2 = .332.

Finally, we analysed the source error data contingent upon the original source of the ideas, rather than the retrieval task (see Figure 1). In order to do this, we took the correct source recall for one set of items, and the wrong-source recall for the complementary set, and calculated the proportion of wrong-source errors for each source. Overall, there was no evidence of a difference in the rate of source errors from each source, (Own responses, M = 0.113, SD = 0.15, Partner responses, M = .169, SD = 0.26), F < 1. Order did not affect source errors, and there was no source by order interaction (both Fs < 1.03, ps > .31).

Experiment 1 therefore produced two clear patterns. Consistent with our expectations, participants made more source-errors in the recall-partner task, suggesting that the mostly commonly studied source error in recall, unconscious plagiarism, is the less frequent error. At the same time, analysis of source-errors contingent upon the original source suggested that own-ideas and partner's ideas are equally susceptible to being output during the wrong task, at least for the subset of items recalled. We reserve discussion of this pattern until we have further replicated it.

Experiments 2a and 2b

Whilst the pattern of findings in Experiment 1 was quite clear, the experiment had a number of sub-optimal features that we sought to improve. Participants in Experiment 1 recalled their own answers for some category cues and their partner's answers to other category cues within a single experimental session, counterbalanced across order. This methodology is not optimal for looking at biases in source attribution because the short retention interval means that rates of wrong-source errors may have been suppressed. Consequently, in Experiment 2a we sought to increase wrong-source errors by increasing the delay prior to test (Brown & Halliday, 1991; Landau & Marsh, 1997, Marsh & Bower, 1993). Also, although the within-subject manipulation of recall cue is statistically sensitive, it raises the possibility of carryover effects caused by the increased salience of the alternate-source cue which might emerge over a delay. To avoid such a possibility, and to make the task more directly comparable to previous demonstrations of unconscious plagiarism, we used a between-subject manipulation of task. That is, participants only attempted a single form of recall-task during this experiment, and so for half the participants (in the recall-own condition), this was effectively the standard Brown and Murphy (1989) procedure. However, this methodology precludes direct analysis of the source errors contingent upon the original source, and so we ran an additional follow-up study (Experiment 2b) in which participants recalled from both sources. We describe this study later.

In Experiment 1 participants were reminded of the number of target responses they were seeking to recall to each cue. Although participants were free to recall however many they wished to report without guess, the prompt that there were 6 target responses per cue may have induced pressure to guess responses, which may explain why such high error rates were observed with a short delay. Consequently, in Experiment 2a we removed this cue by providing participants with a blank response sheet that did not cue the number of potential responses. In this, and subsequent experiments we recorded all duplications at generation, and removed all such items from subsequent analyses. We also removed the time-limit on recall.

At the same time, we changed from an orthographic cue task to a semantic cue task, in which the dyads originally generated members of semantic categories. This task has been used many times in previous research (e.g. Brown & Murphy, 1989; Brown & Halliday, 1991, Marsh & Landau, 1995), and provides an opportunity to replicate the pattern seen in Experiment 1 with a new task. The primary focus of this study is to determine whether participants would once again falsely intrude their own responses when attempting to recall their partner's more often than they would include their partner's responses when recalling their own.

Experiment 2a

Method

Participants: Forty undergraduate students or volunteers participated either for partial course credit or payment of £12 (c. $19.5) for attendance at 3 test sessions. Twenty were tested in the recall-own condition, and 20 in the recall-partner condition.

Procedure: Participants, working in pairs at their own rate, alternated in orally generating responses to semantic category cues, until each had generated 8 exemplars to each of 8 categories (*Articles of clothing, Fruit, Four-footed animals, Articles of furniture, Weather phenomena, Kitchen utensils, Sports, Parts of the human body*). Each participant returned individually twice for separate recall sessions, after 1 and 7 days. Because testing was done individually, participants did not return for testing at precisely the same time of day, but made individual arrangements with the experimenter, who was present for all test sessions.

In Experiment 2a, for each test, participants were presented with a different half of the category cues. That is, no category cue was tested twice, in order to avoid potential carry-over effects across recall-sessions. Participants were instructed to write down only their own or their partner's responses as appropriate, avoiding responses from the wrong source. Thus, for each recall task there was a maximum of 32 items that could potentially be recalled from the correct source, 32 from the wrong source, plus any number of novel intrusions. All recall tasks were self-paced. Duplications during generation were rare and responses associated with such duplications were excluded from the analyses. We scored correct recall, wrong-source recall and intrusions in the same manner as Experiment 1. Unlike in Experiment 1 participants very occasionally generated responses that were not legitimate members of a category, but were nonetheless clearly semantically related (e.g.

“venison” in response to the cue “Four-footed animals”). We did not correct such responses, and nor did we exclude them from the final analyses.

Results and Discussion

Mean frequencies of correct recall, wrong-source errors and intrusions for the recall-own, and recall-partner tasks are shown in Table 2.

Performance (correct recall, wrong-source recall, intrusions) was examined using separate 2 (delay: 1 day vs. 1 week), x 2 (task: recall-own vs. recall-partner) mixed ANOVAs with repeated measures on the first factor. Correct recall decreased over the delay, $F(1,38) = 96.21$, $p < .001$, $MSe = 6.06$, partial $eta^2 = 0.717$, was higher for the recall-own task, $F(1,38) = 77.07$, $p < .001$, $MSe = 14.99$, partial $eta^2 = 0.670$, but there was no interaction, $F < 1$. Wrong-source recall increased after a delay, $F(1,38) = 19.79$, $p < .001$, $MSe = 5.35$, partial $eta^2 = 0.342$, was higher in the recall-partner task, $F(1,38) = 13.41$, $p = .001$, $MSe = 10.87$, partial $eta^2 = 0.261$, but there was no interaction, $F(1,38) = 1.13$, $MSe = 5.35$, $p = .294$, partial $eta^2 = .029$. Intrusions of novel answers also increased after a delay, $F(1,38) = 15.45$, $p < .001$, $MSe = 4.31$, partial $eta^2 = 0.289$, and were more frequent in the recall-partner task, $F(1,38) = 16.38$, $p < .001$, $MSe = 6.32$, partial $eta^2 = 0.301$, but there was no interaction, $F < 1$.

Because recall varied across conditions and delay, we also looked at the output-bound proportion of source errors, as before. There was a higher proportion of wrong-source errors after a delay, $F(1,38) = 24.26$, $p = .001$, $MSe = .010$, partial $eta^2 = .390$, and for the recall-partner task, $F(1,38) = 22.66$, $p < .001$, $MSe = .017$, partial $eta^2 = .374$, but there was no interaction between delay and task, $F(1,30) = 2.15$, $p = .15$, $MSe = .010$, partial $eta^2 = .054$.

Experiment 2a therefore replicated the pattern seen in Experiment 1. Both after a day and a week, recall-partner recall contained twice as many source errors than did recall-own recall. Delay increased wrong-source recall errors overall but did not interact with the nature of the errors. The same pattern was also observed for intrusion errors: participants were biased towards attributing new items to their partners at both delays.

However, as we acknowledged above, restricting participants to only one kind of retrieval task precluded formal analysis of the source-contingent error rates. Nevertheless, the group-mean data shown in Table 2 is consistent with the observation in Experiment 1 that the bias in task-based analyses is absent in the source-based analyses. Based on the group averages, the one-day delay condition showed that own ideas were wrongly attributed 13.6% of the time, and partner ideas 9.0% of the time, and after a week's delay, own ideas were misattributed 27.3% of the time, and partner ideas 24.3%. Because we are not able to statistically confirm this pattern, we ran a follow up study in which participants were asked to recall ideas from both sources.

Experiment 2b

Method

Participants: Twenty participants were recruited from the same volunteer panel, on the same basis as Experiment 2a, but one participant did not attend the final test session, and so did not contribute data to the analyses.

Procedure: The procedure duplicated Experiment 2a in all respects other than the final test. Participants were required to recall ideas from both sources using a response sheet with two columns, headed “Own answers” and “Partner’s answers”. Participants were free to report all the previous responses to each cue that they could recall, but were required to identify the source of each response by means of these two columns.

Results and Discussion

We begin with an analysis of task-contingent performance. Responses in each column were treated equivalently to the single-source recall data in Experiment 2a, that is items written in the “Own answers” were treated as items recalled to the recall-own instruction, and those in the “Partner’s answers” treated as recall-partner items. These data are shown in Table 3.

Participants correctly recalled more own ideas than partner ideas, $F(1,18) = 142.4$, $p < .001$, $MSe = 7.04$, partial $eta^2 = .888$, and more ideas after 1 day than one week, $F(1,18) = 32.25$, $p < .001$, $MSe = 17.65$, partial $eta^2 = .642$, but these two factors did not interact, $F < 1$. The absolute rate of wrong-source errors was higher after a delay, $F(1,18) = 19.23$, $p< .001$, $MSe = 3.65$, partial $eta^2 = .517$, but did not differ by source, $F < 1$, and there was no interaction between source and delay, $F(1,18) = 1.59$, $MSe = 0.953$, $p = .213$, partial $eta^2 = .085$. Source errors as a proportion of output written in a column were higher in the recall-partner column, $F(1,18) = 6.39$, $p = .021$, $MSe = .006$, partial $eta^2 = .262$. The proportion of source errors increased over a delay, $F(1,18) = 25.85$, $p < .001$, $MSe = .006$, partial $eta^2 = .590$ and there was an interaction between source and delay, $F(1,18) = 4.98$, $p = .039$, $MSe = .003$, partial $eta^2 = .217$. Intrusion errors were more common after a delay, $F(1,18) = 20.33$, $p < .001$, $MSe = 4.79$, partial $eta^2 = .530$, and more common in the partner-answer column, $F(1,14) = 16.52$, $p= .001$, $MSe = 2.68$, partial $eta^2 = .479$, but these two factors did not interact, $F < 1$.

The error data were then analyzed contingent upon the original source. Analysis of the proportion of wrong-source errors from each source showed a main effect of delay,

$F(1,18) = 23.18$, $p < .001$, MSe = .011, partial $eta^2 = .563$, a numerical trend for more errors in attributing partner response, $F(1,18) = 3.82$, $p = .066$, MSe = .008, partial $eta^2 = .175$, but no interaction, $F < 1$. After one day 6.3% (SD 5.3%) of generated own ideas were attributed to the wrong source compared to 9.4% (SD 7.3%) of generated partner ideas. After 1 week, these rates increased to 17.2% (SD 11.3%) of own ideas and 22.1% (SD 16.9%) of partner ideas. Thus, the pattern resembled that seen in Experiment 1, with a non-significant tendency for partner ideas to be wrongly attributed more often than own ideas. But, given that this effect was not significant in either experiment, the safest conclusion is that the source-contingent analysis shows no reliable pattern of bias across sources. We return to this issue after we present Experiment 3.

Experiment 2b was different from the previous studies reported here in that it required participants to retrieve previous answers from both sources, and then to attribute each response to a previous source. Thus, throughout the retrieval period, participants were aware of the salience of the source during retrieval, and could not output a response without making an explicit source-judgment. Thus, it seems unlikely that source-neglect (Johnson, Hashtroudi & Lindsay, 1993) contributed to the pattern of errors observed in this study, whilst it could have occurred in previous experiments. This methodological change may have contributed to the failure to replicate the significant effect of retrieval task on the absolute rate of source-errors: here the numerical pattern was consistent with that seen earlier, but it did not reach significance.

Experiment 3

So far, source errors have been analyzed a function of task and source, and the two methods appear to give different patterns. The task-based analyses consistently show more errors in the recall-partner task, while the source-based analyses show little evidence for a difference in rate of errors for own or partner's responses. At first blush, these two patterns may appear incompatible, but in fact both effects may reflect the differential availability of items from each source.

All the experiments to date have shown higher correct recall for own responses than for a partner's responses, in line with the generation effect (Slamecka & Graf, 1978). Thus, at test, regardless of the retrieval task, more own-responses than partner-responses may be available for retrieval. If so then the recall-own task combines highly available correct responses (own-ideas) with less available incorrect responses (partner-ideas) whilst the recall-partner has the reverse combination. In contrast, the source-based analysis corrects for the different base-rates of recall by looking at highly available errors (own ideas recalled incorrectly during the recall partner task) as a proportion of highly available ideas (own ideas recalled in both tasks) and comparing this to the proportion of less available errors (partner ideas recalled incorrectly) as a proportion of less available responses (all partner ideas). Thus, the source-based analysis does not lead to an expectation of bias across source if the two sources differ in availability.

A difficulty with the foregoing argument is that we do not yet have a measure of availability, only what is reported. What is reported during a retrieval task is likely to be a function of the availability (early selection) but also reflects the likelihood of an idea being reported (late correction). The previous experiments do not allow us to separate these processes, and so we do not have a pure measure of the availability of the ideas during each retrieval task. This is the main purpose of Experiment 3.

What we do know with more certainty is that the two retrieval tasks differ in rates of intrusion errors. Given that intrusion errors are not ideas associated with sources that differ in strength, this suggests a role for output monitoring, or late-correction, in the genesis of these errors. One simple late-correction explanation for this is that report criterion differs across the two recall-tasks, such that people are more liberal when trying to report their partner's responses. This would lead to more intrusion errors, but it would also lead to more wrong-source errors being output. Thus, wrong-source errors might involve a late-correction element.

Source-monitoring theory can also be used to explain why late-correction processes may differ across tasks. According to the Source-Monitoring Framework (Johnson, Hashtroudi & Lindsay, 1993), source specifying information can be neglected during demanding retrieval tasks (Marsh, Landau & Hicks, 1997). Given that the evidence suggests that recalling a partner's ideas is more demanding than recalling our own ideas, it is conceivable that source-neglect is more likely to occur during the recall-partner task, resulting in more intrusions, and also more source-errors. Alternately, the pattern of errors could reflect an attributional process that applies in the absence of source-specifying information , the *it-had-to-be-you* heuristic (e.g. Johnson & Raye, 1981; Hoffman, 1997). Rather than failing to consider source, participants may lack the necessary information to accurately judge source, and use this absence of information as being (wrongly) diagnostic of the idea being externally generated. This too would predict more intrusions and source-errors during the recall-partner task.

Experiment 3 was designed to evaluate the role of early selection and late correction processes in the production of wrong-source errors and intrusions. The role of late-correction in the maintenance of memory accuracy has been studied most frequently by contrasting free- and forced-report, a methodology developed by Koriat and Goldsmith to test their memory-control framework (Koriat & Goldsmith, 1994, 1996, 2000; Goldsmith and Koriat, 2008). The essence of this methodology is that participants generate answers to a cue either under forced report instructions, or with the option to withhold a response. The extent to which people are able to trade-off memory quantity (how many correct answers are given) for memory accuracy (what proportion of their answers are correct), and moderate this in response to goals for accuracy or quantity, is a measure of memory control due to late correction.

More recently, the framework has been further developed to acknowledge the potential role of early-selection processes in the genesis of memory errors (e.g. Halamish, Goldsmith & Jacoby, 2012; Goldsmith, 2016; Guzel & Higham, 2013). This work has largely been based upon the use of extended recall (Bousfield & Rosner, 1970; Halamish, Goldsmith & Jacoby, 2012; Kahana, Dolan, Dauder & Wingfield, 2005). In this procedure, participants are asked to report everything that comes to mind in response to a memory cue, but to indicate on-line during retrieval which items are appropriate responses to the retrieval cue. That is, participants are asked to report all the contents of their early selection processes, and then control their output through appropriate late-correction decisions. Kahana et al (2005) used this method to look at source-specified recall in a multi-list experiment in which participants repeatedly studied and were then tested on lists of words. Participants were instructed to recall only the items from the last list studied, but were told to report everything that came to mind during this process. Monitoring was measured by instructing people to manually indicate with a button press every time they reported an item they believed had not come from the last study list. These data were used to determine both the extent to which retrieval was specific to the last list studied, and the ability to monitor the accuracy of ideas that came to mind, and so provides an ideal template for our final experiment.[1]

However, rather than use a button press, we adapted this approach to the present paradigm by asking people to write down all responses that came to mind during retrieval on a response sheet with two headed columns. One column was headed with the retrieval cue (recall own, or recall partner as appropriate), with the other headed “other”, for any other answers that came to mind. Thus, if an answer came to mind that the participant knew to from the wrong source, or that they thought was novel, rather than withhold it as they would have in Experiments 1 and 2, they were encouraged to report it by writing it in the “other” column. We preferred this adaptation of the method over a button press for two reasons. First, it retained the use of written responses we used previously, and second, monitoring accuracy is not confounded with prospective memory. Button presses are an easily omitted response, which could reflect a prospective memory failure as much as a monitoring failure, whereas writing a response down in one of two columns ensures monitoring occurs for every item output.

The extent to which items are available during retrieval is indexed by the tendency for items to be written anywhere on the response sheet. Our expectation was that own-answers would feature as potential competitors during a recall-partner task more often than partner-answers would feature during the recall-own task. Within this set of generated items, we can then examine the accuracy with which participants report items in the correct columns: that is, we can ask what proportion of intrusions and wrong-source errors are

[1] Halamish, Goldsmith and Jacoby (2012 – Experiment 2) also used an extended recall methodology in which participants wrote down all responses that came to mind during retrieval to cue. However, their study materials were paired-associates, and so only a single response to a cue was objectively correct. Our paradigm involved participants retrieving multiple responses to a single (source) cue and so is closer to the approach by Kahana *et al*., (2005).

written in the target column (rather than the “other” column) separately for the recall-own and recall-partner tasks. If participants show a bias towards attributing weaker memories to their partners, we would expect a higher proportion of responses generated at retrieval to be mistakenly included in the recall-partner task than in the recall-own task.

Experiment 3 therefore replicates Experiment 2a, except that we used the extended recall version of the retrieval task. Additionally, we replaced our within-subjects manipulation of delay with a between-subjects manipulation, so that participants were only required for a single test session. An account of the previous bias in wrong-source errors based entirely upon early-selection would suggest that own responses would be generated more often (during the recall-partner task) than partner responses would be generated (during the recall-own task), without any bias in subsequent monitoring processes. In contrast, an account based entirely upon late-selection would be that the availability of responses from both sources is equivalent, but output monitoring is less stringent during the recall-partner task, leading to more own-responses being reported. We think such an account is unlikely, because the base-rate of generation of the two sources is not matched, but we do not rule out a role of bias in late-correction in conjunction with an early-selection bias. This view would be supported by the greater propensity to generate own responses, and the greater propensity to accept them given that they are generated.

This procedure also allows us to distinguish whether the increase in wrong-source recall that was observed with delay can be attributed to early selection or late correction processes. With increased delay, early selection processes may be less able to selectively generate target items rather than non-target items. If so, then wrong-source errors would constitute a higher proportion of generated items, and so be more likely to intrude during recall, even as overall recall drops. Alternately, late-correction processes could contribute if the ability to distinguish source drops with delay: even if no more wrong-source errors are generated, more might be accepted as task compliant. Given that increased source errors are found with increased delay in recognition paradigms (Bornstein & Lecompte, 1995), we strongly suspect that late-correction processes are partially responsible for the increase in the report of wrong-source errors with a delay, but we cannot rule out a role for early selection processes as well.

The inclusion of the extended recall methodology provides a particular test of the use of the *it-had-to-be-you* heuristic. This predicts that during standard recall procedure, intrusion errors are likely to be rejected during a recall-own task, but accepted during a recall-partner task. If this is so, then in an extended recall task, the non-target column should include more intrusions in the recall-own task (because the responses are rejected as not being own) than in the recall-partner task (because they have been accepted as task compliant).

Method

Participants: Eighty-nine undergraduate students or volunteers attended for partial course credit or payment of £8 for attendance. However, only 81 attended the second session of the study. In the recall-own condition 20 participants were tested after a 1 day delay, and 20 after a 1 week delay. In the recall-partner condition, 20 were tested after 1 day, and 21 after 1 week.

Procedure: This followed Experiment 2a closely, except for the final test phase. Participants returned for only a single test phase, either after 1 day or 1 week, and were tested on their recall for all categories at that delay. This effectively doubled the potential number of items that could be recalled in a condition, relative to Experiment 2a. The second change was the use of an extended-recall task. In each condition (recall-own or recall-partner), participants were given a series of response sheets headed by a category cue (e.g. “items of clothing”), and two columns for responses. One column was for task-compliant responses (i.e. own answers in the recall-own task, partner answers in the recall-partner task), and the other column was for anything else that came to mind. Participants were given the following instructions as to how the sheet should be filled out:

“I want you to try and remember what {you / your partner} came up with last week for each category. As you try to remember {your own/ your partner’s} ideas for each category, you might remember some ideas {your partner / you} came up with and some entirely new, random ideas that you know neither of you actually said. Whenever that happens I want you to write down those ideas along the side of the sheet. You can write down as many or as few ideas along the side that you know aren’t {yours / your partners} as you like or as come to mind, so that the sheet could look something like this: [Show what filled-out recall sheets look like.] So essentially, try and write out everything that comes to mind. Whenever it’s {your own / your partner’s} idea write it down here, whenever it’s {your partner’s / your own} idea or an entirely new idea write it down here, but the task is to try and remember {your own / your partner’s} examples. I don’t want you to search your memory for {your partner’s / your own} examples or new ideas, just write them down if they come to mind as you try and remember what {you / your partner} came up with.”

Results and Discussion

Table 4 reports the same measures that were reported in the first two experiments. However, because of the use of extended recall, it includes both the number of responses that were included as task compliant (i.e. reported) and the total number of responses of each kind that were generated (i.e. written anywhere on the response sheet, reported and excluded). Before we present the data, we need to explain our terminology. Participants were instructed to recall the responses associated with one source, and so for the responses written in the task-compliant column, these could be classified as correct, wrong-source errors or intrusions. However, in the "other column", participants also wrote examples of the same three kinds of response, namely their own original responses, their partner's original responses and new responses. For the old ideas it is ambiguous whether these responses are errors or not. For example, when attempting to recall a partner's response, generating a response that was originally self-generated is an error, but withholding it is correct. Conversely, generating a partner's response is correct, but withholding it would be incorrect. The same issues apply to own ideas and intrusions, except in the latter case, withholding is always a correct response.

Consequently, in addition to the analyses of reported outputs, we looked at the generated data in two ways. First, we looked at all the ideas generated in response to the retrieval goals, regardless of whether they were reported or not. This enabled us to classify ideas that came to mind as correct, wrong-source errors or intrusions, using a task-contingent analysis, just as for those that were ultimately reported. We then looked at the proportion of the generated ideas that were reported as task-compliant. That is, we looked at the ability to monitor the accuracy of available ideas, independent of the frequency with which each type of idea came to mind.

The first question of interest is whether using the extended-recall paradigm alters the pattern of responses ultimately reported. Consequently, we first analyzed the responses written in the target columns, using a task-contingent analysis, to determine whether we replicated the pattern reported in Experiments 1 and 2. These data are reported in the top half of Table 4.

For correct recall, recall was higher for own answers than for partner answers, $F(1,77) = 62.18$, MSe = 45.19, $p < .001$, partial $eta^2 = .447$, was higher with a shorter delay, $F(1,61) = 51.11$, MSe = 45.19, $p < .001$, partial $eta^2 = .399$, but these two factors did not interact, $F < 1$. For wrong-source errors, more errors occurred in the recall-partner task than in the recall-own task, $F(1,77) = 14.11$, MSe = 17.01, $p <.001$, partial $eta^2 = .155$, and following a longer rather than a shorter delay, $F(1,77) = 43.46$, MSe = 17.01, $p < .001$, partial $eta^2 = .361$, but these two factors did not interact, $F < 1$. More intrusion errors were reported during the recall-partner task than during the recall-own task, $F(1,77) = 23.80$, MSe = 20.05, $p < .001$, partial $eta^2 = .236$, and after a longer delay, $F(1,77) = 30.59$, MSe = 20.05, $p < .001$, partial $eta^2 = .284$, but these two factors did not interact, $F(1,61) = 1.12$, MSe = 20.05, $p = .293$, partial $eta^2 = .014$.

Because the number of responses included varied across tasks and delay, we also looked at wrong-source errors as a proportion of responses output. As before, wrong-source errors constituted a higher proportion of answers output during the recall-partner task, $F(1,77) = 24.28$, MSe = .005, $p < .001$, partial $eta^2 = .240$, and after a delay, $F(1,77) = 50.14$, MSe = .005, $p < .001$, partial $eta^2 = .394$, but there was no interaction, $F < 1$.

Experiment 3 resembled Experiment 2a in that participants only completed a single retrieval task and so it was not possible to carry out an analysis of source errors contingent upon source. Looking at the group means in Table 4 suggests that there was no clear evidence of a bias in errors from each source. After 1 day, 10.5% of own ideas and 7.1% of partner ideas were mistakenly reported, while after 1 week, 24.8% of own ideas and 24.0% of partner ideas were mistakenly reported[2].

[2] Note: this analysis assumes that these two responses are exclusive, but this may not be the case. For example, participants who wrongly reported an own-item during the recall-partner task may not have excluded it during the recall-own task. The assumption of exclusivity is challenged by changes in

Thus, in terms of the answers reported as task-compliant, all previous patterns were replicated, suggesting that the use of the recall-everything instructions did not fundamentally alter participants' memory report. Participants were better able to correctly recall their own responses than their partner's, and more likely to wrongly include their own responses than their partner's. This was true whether expressed as an absolute number of errors, or as a proportion of answers output. In addition, as in Experiment 2a, we found more wrong-source errors after a 1-week delay than after a 1-day delay. However, when we looked at the source-errors contingent upon the original source, there was no apparent tendency to misattribute one source more than another. Having established that the recall-everything methodology produces equivalent performance to that seen previously, our attention now turns to examining the generated ideas to determine the roles of early selection and late correction in the generation of recall-errors and intrusions.

Early selection processes

The next analysis looked at the total number of answers that were written down by the participant anywhere on the response sheet, ignoring the nature of those answers (i.e. correct-source, wrong-source or intrusions), which is the sum of the data shown in the bottom half of Table 4. Surprisingly, participants wrote down as many responses for the recall-partner task (M= 93.34, SD = 22.43) as they did the recall-own task (M= 95.90, SD = 19.72), F < 1, and as many responses after 1 week (M = 95.24, SD = 24.63) as 1 day (M =93.95, SD = 16.88), F < 1, with no interaction between these two factors, F < 1. Thus neither task nor source appeared to influence participant's willingness to respond. Of course, this does not mean that the same kinds of answers came to mind across these different conditions.

---

report threshold across tasks, or from source-neglect which would result in participants reporting an item regardless of the goals of the task.

The availability of responses from different sources (i.e. excluding intrusion errors) was explored with a 2 (source: own vs. partner) x 2 (task: recall own vs. recall partner) x 2 (delay: 1 day vs. 1 week) mixed ANOVA on frequency of responses with repeated measures on the first factor. Overall, participants generated more of their own answers than their partner's, $F(1,77) = 145.02$, $MSe = 37.05$, $p < .001$, partial $eta^2 = .653$, which was moderated by an interaction with task, $F(1,77) = 36.19$, $MSe = 37.05$, $p < .001$, partial $eta^2 = 0.320$. Follow-up tests indicated that for the recall-own task, participants generated more of their own responses (M = 49.05, SD = 9.40) than their partner's responses (M = 31.78, SD = 7.89), $F(1,77) = 161.1$, $p < .001$, partial $eta^2 = .677$, whilst for the recall-partner task participants generated fewer of their partner's answers (M = 35.39, SD = 7.79) than their own (M = 41.16, SD = 9.42). Consistent with the successful operation of early-selection processes, follow-up tests also indicated that more own answers were generated during the recall-own task than during the recall-partner task, $F(1,77) = 13.85$, $p = .003$ $MSe = 90.98$, partial $eta^2 = .152$ and more partner-answers were generated during the recall-partner task than during the recall-own task, $F(1,77) = 4.36$, $p = .04$, $MSe = 60.68$, partial $eta^2 = .054$.

There was no overall effect of task on number of responses written, $F(1,77) = 1.613$, $p = .208$, $MSe = 114.6$, partial $eta^2 = .021$, but there was an overall delay effect, such that more old responses were generated after 1 day (M = 42.09, SD = 7.57) than after 1 week (M = 36.60, SD = 7.57), $F(1,77) = 10.63$, $MSe = 114.6$, $p = .002$, partial $eta^2 =.121$. However, there were no other significant higher order interactions, $F < 1$ in all cases.

The emergence of a delay effect in recall of old responses, coupled with the absence of such an effect in overall number of responses is reconciled by inspecting the rate of intrusion errors. Overall, participants were more likely to include a novel response in their output after 1 week than after 1 day, $F(1,77) = 24.24$, $MSe = 126.05$, $p < .001$, partial eta^2

= .239. However recall task made no impact upon rate of intrusions that were generated, $F < 1$, and nor did this factor interact with delay, $F < 1$.

Thus, whilst the total number of responses generated at retrieval did not vary across task or delay, the nature of those responses did. Expressed as a proportion of items generated, items from the non-target source were generated more often during the recall-partner task than the recall-own task, $F(1,77) = 44.20$, $MSe = .005$, $p < .001$, partial $eta^2 = .365$, but less often after a delay, $F(1,61) = 7.07$, $MSe = .005$, $p = .01$, partial $eta^2 = .084$. There was no task by delay interaction on the proportion of wrong-source errors generated, $F(1,61) = 1.046$, $MSe = .005$, $p = .310$, partial $eta^2 = .013$.

Thus different patterns emerge for the generation of intrusions, and wrong-source errors during retrieval, as is illustrated in the panel A of Figure 2. The generation of intrusion errors becomes more likely after a delay, but is uninfluenced by the retrieval task, whilst the generation of wrong-source errors is influenced by task, but not delay.

Late correction processes

The final analysis was designed to test the effects of task and delay on late-correction processes. We looked at the ability to correctly reject wrong-source and intrusion errors by comparing the proportions of wrong-source errors and intrusions that were wrongly assigned to the task-compliant column at each delay. These proportions are illustrated in the panel B of Figure 2. We ran a 2 (error type: wrong-source errors vs. intrusion errors) x 2 (task: recall-own vs. recall-partner) x 2 (delay: 1 day vs. 1 week) mixed ANOVA on the proportion of errors reported with repeated measures on the first factor. Participants were more likely to report an intrusion error than a wrong-source error, $F(1,77) = 42.67$, $MSe = .031$, $p < .001$, partial $eta^2 = .357$, included more errors in the recall-partner task than the recall-own task, $F(1,77) = 12.07$, $MSe = .061$, $p = .001$, partial $eta^2 = .135$, and included more errors after a delay, $F(1,77) = 11.40$, $MSe = .061$, $p = .001$, partial $eta^2 = .129$. The only significant higher-order interaction was between error type and task, $F(1,77) = 5.88$, $MSe = .031$, $p = .018$, partial $eta^2 = .071$. Follow up tests revealed that a higher proportion of intrusions were accepted for the recall-partner task than the recall-own task, $F(1,77) = 12.63$, $p = .001$, but the rate of acceptance of wrong-source errors did not differ across tasks, $F(1,77) = 3.47$, $p = .066$. No other interactions were significant, $F(1,77) < 3.28$, $p > .073$ in all cases.

Thus in summary, although the pattern of performance for reported items across retrieval task and delay is the same for intrusions and wrong-source errors, it is clear that their genesis is different (see Table 4, and panel C of Figure 2 for illustrative purposes). Intrusion errors were generated at equal rates across retrieval tasks whilst source errors showed an effect of retrieval task. That is, there was an early-selection effect on wrong-source errors that was absent for intrusion errors. Additionally, it is clear that generated wrong-source errors were more likely to be rejected than generated intrusions, suggesting active late-correction processes.

An obvious potential critique of this experiment is that the addition of the “other” column, coupled with the instructions to report everything, has fundamentally changed the task. Clearly, our ability to interpret the data from Experiment 3 rests upon people accurately reporting what comes to mind, without altering the way in which they sought to recall the target answers. Whilst we cannot absolutely refute this critique, we believe there are good grounds for drawing firm conclusions from this experiment.

The main reason for our confidence is that our findings entirely replicate Experiment 2a, apart from some clearly predicted deviations. With regards to the memories correctly reported as task-compliant, the pattern in Experiment 3 is strikingly similar to Experiment 2a, suggesting that the use of extended recall had not fundamentally altered performance. In particular, source-errors on the recall-own task were highly consistent across Experiments: Experiment 3 found error rates of 5.2% (after 1 day) and 16.0% (after 1 week), which resemble those seen in Experiment 2a (5.7% and 13.4% respectively). For errors on the recall-partner task, the pattern was also similar, though errors were slightly lower in Experiment 3 (1 day, 12.7%, 1 week, 23.7%) than seen in Experiment 2a (1 day 16.3%, 1 week, 30.6%).

With regards to intrusion errors, the pattern was entirely as we predicted from use of the *it-had-to-be-you* heuristic. This heuristic applies when people bring an idea to mind, and then decide whether or not to report it. This predicts that intrusion errors should appear in the non-target column for the recall-own task, but in the target column for the recall-partner task. This is precisely what was found. Based on group averages, Experiment 2a showed that after a day intrusions constituted 3.0% of recall-own output, and 15.6% of recall-partner output, and after a week, the equivalent figures were 10.3% and 22.6%. The task-compliant data in Experiment 3 were remarkably similar. After a day, intrusions constituted 2.7% of recall-own, and 11.2% of recall-partner output, and these figures rose to 11.5% and 23.5% after a week. However, additionally, intrusion errors showed the pattern expected in the full set of items generated (i.e. written in either column), with little difference in intrusion rates across recall tasks. After one day, the intrusions rates were 9.0% and 11.9% for recall-own and recall-partner tasks respectively, which rose to 22.4% and 23.9% after a week. Thus, as we predicted, for the recall-own task participants wrote down intrusions in the “other” column that they would otherwise have withheld, whilst in the recall-partner task they were already including them in the task-compliant column.

The major focus of Experiment 3 was to understand the genesis of wrong-source errors, and to explore whether retrieval-task and delay influence wrong-source errors in different ways. There was clear evidence that this was the case, as consideration of the 3 panels of Figure 2 reveals. The nature of the retrieval task had no impact upon the number of intrusions generated but it did influence the number of source errors generated. Conversely, delay had no impact upon the number of wrong-source errors generated, but it increased the number of intrusions generated. However, this pattern of response generation did not map directly onto reported performance, because the propensity to accept a generated item as being task compliant itself differed across conditions.

Overall, participants were more likely to include generated intrusion errors into their task-compliant responses than old-answers from the wrong source. This is not surprising, because old responses that came to mind will have been accompanied by source-specifying information that allowed participants to reject them as not being task compliant. Additionally, participants were more liberal in accepting generated responses as task compliant during the recall-partner task. The fact that this effect was stronger for intrusion errors than wrong-source errors is also consistent with the idea that the wrong-source errors were accompanied by source-specifying information that helped counteract the tendency to report their own answers as being their partner's (and to a lesser extent their partner's responses as their own, although the key interaction was not significant). This willingness to accept a generated answer as being task compliant increased after a delay, perhaps as a result of a lower report threshold, or a reduction in source-specifying details available to counteract this tendency. Thus, whilst the observed pattern shown in panel C appears to suggest that intrusions and wrong-source errors are both equally frequent, and equally influenced by task and delay, it is clear from the different patterns for intrusion errors and wrong-source errors that there is a strong role for late-correction-based monitoring in the production of the previously observed bias in wrong-source errors.

At the same time the data on generated answers in Table 4 demonstrate an early selection effect, because responses associated with a particular source were generated more often when that source was the target of recall than they were when the other source was the target of recall. It is possible that participants may have not written down all the items that they thought of, and so this apparent early-selection effect could instead be thought of as a covert late-correction effect. Whilst this cannot be ruled out, it is not clear why this pattern did not vary with delay. Delay led to a doubling of intrusion errors being written, consistent with a more liberal criterion for report. But, delay led to a reduction in wrong source-errors being reported, and no change in magnitude of the early selection effect observed. If participants were withholding non-target responses, one might expect this effect to mirror that seen for other non-target responses – intrusion errors – but this was not the case.

Although we did see an early selection effect, it was relatively small. A potential reason for this is that the retrieval cue we specified is only one means of solving the task we set participants. Our instructions stressed one potential retrieval cue – one person in the dyad – but participants may have used other cues to retrieval, such as the semantic category cue, the environmental context, and so forth. Such retrieval cues are likely to access both target and non-target items, requiring late correction processes to achieve accurate recall. Thus, in our experiment, the observed early selection effect may have been small because other retrieval cues were available. Whilst this idea is speculative, our belief is that both early selection and late correction processes are likely to be involved in many retrieval tasks, but the weighting of each will reflect the ability of available retrieval cues to precisely access target memories.

In summary, it is clear that both early-selection and late-correction processes play a role in the reported rates of wrong-source errors seen in Experiment 2a, replicated in the reported responses in Experiment 3. As predicted, our own responses are more likely to intrude as competitors during a recall-partner task than our partner's answers are to intrude during a recall-own task. Additionally, we are more liberal in accepting generated responses as being task compliant for a recall-partner task, and following a delay. These effects are additive on the wrong-source errors we report at output, and so we are particularly prone to mistakenly give away our own responses after a delay.

General Discussion

Four experiments explored the extent to which source-errors occur during free recall. Previously, there has been an extensive literature demonstrating that people are likely to mistakenly include other's ideas as their own when trying to recall their own idea, even when trying not to, a phenomenon known as unconscious plagiarism. What has not been studied previously is the tendency to mistakenly report own ideas as having come from a partner or the tendency to mistakenly withhold answers from the correct source. Both of these findings were evident in the present set of studies.

Across all four experiments, participants were more likely to mistakenly report one of their own ideas when recalling their partner's idea than they were to make a standard unconscious plagiarism error. This finding was demonstrated in all experiments when measured as a proportion of response output, and in three out of four experiments when measured in absolute terms. Thus, one product of the current set of studies is to act as a counterweight to the large number of studies that have focused on unconscious plagiarism errors only. Whilst unconscious plagiarism errors resemble the egocentric biases reported in evaluations of contributions to joint work (e.g. Ross & Sicoly, 1979), the current studies suggest that unconscious plagiarism should not be interpreted as a self-serving bias. Rather, such errors are better understood as one particular instantiation of the broad problem of source-specified retrieval.

A second outcome of the current set of studies is to highlight the role of generation in the production of source-errors in recall, a topic that has been relatively neglected in the discussion of unconscious plagiarism to date. Instead theorists have sought to explain unconscious plagiarism in terms of memory strength (which we discuss below), or the source monitoring framework. However, neglecting the role of generation in free recall undermines attempts to draw theoretical conclusions about source monitoring. For example consider word-frequency as a manipulation. Studies of source-monitoring have shown more accurate source judgments for low-frequency words than for high-frequency words (Johnson, Raye, Foley & Foley, 1981). However, Marsh and Landau (1995) found no word-frequency effect on rates of plagiarism in their recall-own tasks, and so concluded that plagiarism errors are not due to source-monitoring, but occur as a result of mistaken fluency. The present results generated by the extended recall paradigm suggest an alternative perspective. What Marsh and Landau (1995) failed to acknowledge is that a recognition-based source-monitoring paradigm automatically controls for the generation of different kinds of ideas by providing them experimentally. Given an equal probability of being judged, source-monitoring accuracy is higher for low-frequency items (Johnson et al, 1981). But in a recall-task, the likelihood of generating high and low-frequency items may not been matched: participants in the Marsh and Landau (1995) study are likely to have generated more high-frequency exemplars, even if they reported no more high-frequency responses in their output. That is, a bias in early selection could be balanced against source-monitoring based late-correction processes to produce a null effect on plagiarism rates. Thus, the conclusion that source-monitoring is not involved in unconscious plagiarism was premature.

Up until this point, we have discussed the patterns of findings for the experiments in terms of the source-monitoring framework, and early-selection and late-correction processes. However, there is a strength-based signal detection based account of unconscious memory that has been used to explain unconscious plagiarism findings. This can explain some, but not all, of the data presented here. Figure 3 depicts Marsh and Bower's (1993) model in which, the strength of responses that are initially self-generated is higher on average than the strength of responses generated by a partner, which are in turn higher than new ideas. At retrieval, responses that are generated with sufficient strength to pass the lower criterion are judged as old, whilst those with enough strength to pass the upper criterion are judged as being self-generated. Thus, answers are judged as coming from a partner if they are generated with sufficient strength to be recognized as old, but with insufficient strength to be judged as one's own.

This relative-strength model was first proposed by Marsh and Bower (1993) and later tested by Marsh and Landau (1995). To test the model's assumptions, Marsh and Landau (1995) had participants engage in a lexical decision task that involved items either previously generated or heard by the participant. Consistent with the model, words that had been previously generated were processed more quickly in the lexical decision task than words that had been heard, and words that were (later) plagiarized were processed more quickly than those that were not. Additionally, they showed that ideas later plagiarized in the recall-own task were processed more quickly than those later plagiarized in the generate-new task, consistent with the assumption that the former fall above the criterion for recognition, and the latter fall below it.

Deriving precise predictions from this model regarding the relative frequency of recall-own and recall-partner plagiarism is problematic, because it depends upon the relative strengths of the own- and partner-distributions, and the position of the two criteria. Nevertheless, as illustrated in Figure 3, it is easy to see how the greater propensity to give away ideas than to steal them could emerge. If the upper threshold is set relatively cautiously, this will have the simultaneous effects of reducing the propensity to plagiarize and make intrusion errors during the recall-own task, but to simultaneously increase the propensity to attribute one's own ideas to a partner and to make intrusion errors.

The model is also able to explain the different pattern we observed in the source errors depending upon the different contingency analyses. The starting point for this explanation is the observation that not all available items are recalled. For the sake of the argument we will consider items below the C_{old} criterion to be unavailable. In this instance, performance can only be optimized by setting the C_{own} criterion to judge the source of each old idea that comes to mind, by outputting items above the criterion in the recall-own task and below it in the recall-partner task. The final placement of this criterion will in part rest upon the relative value placed upon each kind of source error, but the numerical impact of different criterion setting can be estimated.

If participants had to distinguish all their own responses from all their partner's responses (as in a source-monitoring task), the optimal performance would be when the C_{own} criterion is set at the intersection of the own and partner distributions. However, in recall, this would predict a small difference in the absolute rate of source errors across source because all partner ideas above the criterion would be reported, but own ideas that fell below the C_{own} criterion would only be reported if they were also above the C_{old} criterion. At the same time, ignoring intrusions, the proportion of source errors by task would be higher for the recall-partner task because fewer correct items would be output. This is what was observed. However, this ignores intrusion errors, which participants cannot.

Placing the C_{own} criterion at the intersection of the own and partner response distributions is non-optimal because it fails to control for errors associated with intrusions: Lowering the criterion below the intersection between the own and partner distributions would be sub-optimal, because although correct recall-own recall would increase in line with the decrease in recall-partner recall, intrusions would also increase. Conversely, raising the C_{own} criterion would benefit overall performance for the opposite reason. The change in the accuracy for the old ideas would be matched, but intrusions would decrease, at least up until the point at which the decrease in intrusions and wrong-source errors for partner recall matched the decrease in correct recall. Clearly, a very conservative C_{own} criterion would remove all errors, but at great cost to correct recall of own ideas. Thus, optimal performance is achieved with a criterion that is above the intersection of the own and partner distributions, but the exact placement of this criterion depends upon the relative strengths of the three distributions (and the value-weighting of each kind of outcome).

The consequence of this criterion is exactly the pattern we observed: as a function of task, a conservative upper criterion is likely to produce relatively few wrong-source errors and intrusions in the recall-own task compared to the recall-partner task. This would be true both in terms of absolute rate, and proportion of output in a task and so it would predict the task-based biases observed here. However, it need not produce any such bias in the source-based contingency analyses. This follows because the source-based analyses only consider responses above the C_{old} criterion, and there are more own-responses than partner responses that pass this criterion. The exact proportions of own and partner ideas that are attributed to the correct source will depend jointly on the placement of the C_{old} and C_{own} criteria, and so no clear bias in source errors is necessarily predicted.

Whilst this model is able to account for these aspects of our data, it struggles to explain other aspects of our data. One is the difference in reporting rates for intrusion errors and wrong-source errors in Experiment 3. In this relative-strength model, the only difference between new items and a partner's answers is their relative strength: against any criterion, during the recall-own task, more wrong-source errors should be generated than intrusions. This was indeed the case, as panel A of Figure 2 illustrates. However, as Panel B illustrates, participants in the recall-own task were much more likely to include generated intrusion errors than generated wrong-source errors. That is, a higher proportion of the generated ideas associated with the weaker source were included at output, which contradicts the model.

The model also has no mechanism by which an early-selection process can occur because the lower threshold provides the criterion for generating a response, whilst the upper criterion provides the means for making a source-judgment. It follows from this observation that the number of own responses correctly reported as own during a recall-own task should match the number of own-responses correctly rejected in the recall-partner task, with an equivalent prediction for partner-responses. But as we saw, participants demonstrated early selection effects, such that they were more likely to generate an idea that was task compliant. One potential criticism of this argument is that participants may not have reported all the responses that they generated. However, in order for there to have been fewer partner-responses generated during a recall-own task than a recall-own task, a shift in criterion across tasks would be required, which is contradicted by the lack of a task effect on intrusion rates.

To summarise, neither the strength-based model, nor a purely post-retrieval source-monitoring account can explain the full pattern of data reported here. The single-strength model fails because it reduces strength to a property of the item, rather than a property of the cue-item match that varies across retrieval tasks, and the post-retrieval source-monitoring account fails because it neglects the role of early selection processes in driving the kinds of items that need to be monitored.

We believe that our focus on early-selection and late-correction processes offers a fruitful direction for further research into how people control the accuracy of their recall (see Goldsmith, 2016 for a similar conclusion). The majority of research in source-memory has been conducted using recognition based procedures, and source-contingent analyses. However, we often need to regulate the accuracy of what we recall during conversation: to date this issue has largely been researched with a recall-based paradigm focusing only on our ability to recall our own previous responses. We believe that the approach developed here, including the use of other forms of recall task, including recalling a partner's ideas, and using an extended recall task, offers the potential to provide further insight into how people control the accuracy of their report, and what factors influence this ability.

References

Bornstein, B. H. & Lecompte, D. C. (1995). A comparison of item and source forgetting. *Psychonomic Bulletin and Review, 2*, 254-259. doi: 10.3758/BF03210966

Bousfield, W. A. & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science, 20*, 75-76. doi: 10.3758/BF03335608

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

Brown, A. S., & Halliday, H. E. (1991). Cryptomnesia and source memory difficulties. *American Journal of Psychology, 104*, 475-490. doi: 10.2307/1422937

Defeldre, A-C. (2005). Inadvertent plagiarism in everyday life. *Applied Cognitive Psychology, 19*, 1-9. doi: 10.1002/acp.1129

Foley, M. A., Foley, H. J., Durley, J. R. & Maitner, A. T. (2006). Anticipating partner's responses: examining item and source memory following interactive exchanges. *Memory & Cognition, 34*, 1539-1547. doi: 10.3758/BF03195917

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. doi: 10.1080/20445911.2013.841674

Goldsmith, M. (2016). Metacognitive quality-control processes in memory retrieval and reporting. In John Dunlosky and Sara K. Tauber (Eds.) The Oxford Handbook of Metamemory (pp xx-xx), Oxford, Oxford University Press. doi: 10.1093/oxfordhb/9780199336746.013.28

Goldsmith, M. & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin and B. Ross (Eds.), *Psychology of learning and motivation, Vol. 48: Memory use as skilled cognition* (pp. 1-60). San Diego, CA: Elsevier.

Guzel, M. A. & Higham, P. A. (2013). Dissociating early- and late-selection processes in recall: the mixed blessing of categorized study lists. *Memory & Cognition, 41*, 683-697. doi: 10.3758/s13421-012-0292-3.

Halamish, V., Goldsmith, M. & Jacoby, L. L. (2012). Source-constrained recall: front-end and back-end control of retrieval quality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 1-15. doi: 10.1037/a0025053

Hashtroudi, S., Johnson, M. K. & Chrosniak, D. (1989). Aging and source monitoring. *Psychology and Aging, 4*, 106-112. doi: 10.1037//0882-7974.4.1.106

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

Hunt, R. & Smith, R. (2014). How distinctive processing enhances hits and reduces false alarms. *Journal of Memory and Language, 75*, 45-57. doi: 10.1016/j.jml.2014.04.007

Hyman, I. E. Jr., Roundhill, R. F., Werner, K. M. & Rabiroff, C. A. (2014). Collaboration inflation: egocentric source monitoring errors following collaborative remembering. *Journal of Applied Research in Memory and Cognition, 3*, 293-299. doi: 10.1016/j.jarmac.2014.04.004

Jacoby, L. J., Kelley, C. M. & McElree, B. (1999). The role of cognitive control: Early selection and late correction. In S. Chaiken & Y. Trope (Eds.) *Dual Process Theories in Social Psychology* (pp. 383-400). New York: Guilford.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117*, 390-394. doi: 10.1037/0096-3445.117.4.390

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

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85. doi: 10.1037/0033-295X.88.1.67

Johnson, M. K., Raye. C. L., Foley, H. J. & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology, 94*, 37-64. doi: 10.2307/1422342

Kahana, M. J., Dolan, E. D., Sauder, C. L. and Wingfield, A. (2005). Intrusions in episodic recall: age differences in editing of overt responses. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*, P92-P97. doi: 10.1093/geronb/60.2.P92

Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition, 17*, 563-571. doi: 10.3758/BF03197079

Koriat, A. & Goldsmith, M. (1994).Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297-315. doi: 10.1037/0096-3445.123.3.297

Koriat, A. & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297-315. Doi: 10.1037/0096-3445.123.3.297

Koriat, A. & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517. doi: 10.1037/0033-295X.103.3.490

Koriat, A. & Goldsmith, M. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481-537. doi: 10.1146/annurev.psych.51.1.481

Landau, J. D. & Marsh, R. L. (1997). Monitoring source in an unconscious plagiarism paradigm. *Psychonomic Bulletin and Review, 4*, 265-270. doi: 10.3758/BF03209404

Macrae, C. N., Bodenhausen, G. V., & Calvini, G. (1999). Contexts of cryptomnesia: May the source be with you. *Social Cognition, 17*, 273-297. doi: 10.1521/soco.1999.17.3.273

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

Marsh, R. L. & Landau, J. D. (1995). Item availability in cryptomnesia: assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1568-1582. doi: 10.1037/0278-7393.21.6.1568

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 886-897. doi: 10.1037/0278-7393.23.4.886

Perfect, T. J., Defeldre, A-C, Elliman, R. & Dehon, H. (2011). No evidence of age-related increases in unconscious plagiarism during free recall. *Memory, 19*, 514-528. doi: 10.1080/09658211.2011.590503

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 and R. Bjork (Eds.). *A Handbook of Memory and Metamemory*, (pp. 285-314). New York: LEA.

Perfect, T. J. & Stark, L-J. (2012). Unconscious plagiarism in recall: attribution to the self, but not for self-relevant reasons. *Europe's Journal of Psychology, 8*, 275-283. doi: 10.5964/ejop.v8i2.459

Ross, M. & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37*, 322-336. doi: 10.1037/0022-3514.37.3.322

Slamecka, N. J. & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592-604. doi: 10.1037/0278-7393.4.6.592

Stark, L-J. & Perfect, T. J. (2006). Elaboration inflation: how your answers become mine. *Applied Cognitive Psychology, 20*, 641-648. doi: 10.1002/acp.1216

Stark, L-J. & Perfect, T. J. (2007). Whose idea was that? Source monitoring for idea-ownership following elaboration. *Memory*, *15,* 776-783. **Doi:** 10.1080/09658210701643042

Stark, L-J. & Perfect, T. J. (2008). The effects of repeated idea elaboration on unconscious plagiarism. *Memory & Cognition, 36*, 65-73. doi: 10.3758/MC.36.1.65

Stark, L-J., Perfect, T. J. & Newstead, S. (2005). When elaboration leads to appropriation: unconscious plagiarism in a creative task, *Memory, 13*, 561-573. doi: 10.1080/09658210444000232

Starns, J. J., Hicks, J. L., Brown, N. L. & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition, 36,* 1-8. doi: 10.3758/MC.36.1.1

Stephenson, B. & Wicklund, R. A. (1983). Self-directed attention and taking the other's perspectives. Journal of Experimental Social Psychology, 19, 58-77. doi: 10.1016/0022-1031(83)90005-7

Taylor, F. K. (1965). Cryptomnesia and plagiarism. *British Journal of Psychiatry, 111*, 1111-1118. doi: 10.1192/bjp.111.480.1111

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

Wicklund, R. A., Reuter, T. & Schiffmann, R. (1988). Acting on ideas: Appropriation to the self. *Basic and Applied Social Psychology, 9*, 13-31. doi: 10.1207/s15324834basp0901_2

Table 1: Experiment 1: The frequency of correct-source recall, wrong-source recall and intrusion errors in recall for recall-own and recall-partner tasks.

	Recall-own		Recall-partner	
	Mean	SD	Mean	SD
Correct source	10.50	3.39	4.95	3.01
Wrong source	0.95	1.74	1.63	1.96
Intrusions	1.40	1.76	2.13	2.62
% Source Error	6.5%	11.4%	17.4%	18.7%

Table 2: Experiment 2a: The frequency of correct-source recall, wrong-source recall and intrusion errors in recall for recall-own, and recall-partner tasks across delay.

	1 day				1 week			
	Recall-Own		Recall-Partner		Recall-Own		Recall-Partner	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Correct Source	23.1	2.19	15.1	3.62	17.3	3.74	10.1	3.19
Wrong Source	1.50	1.43	3.65	2.97	3.25	3.09	6.50	3.45
Intrusions	0.75	1.29	2.80	1.47	2.35	2.11	4.85	3.60
% Source Error	5.7%	5.2%	16.3%	10.4%	13.4%	11.9%	30.6%	16.3%

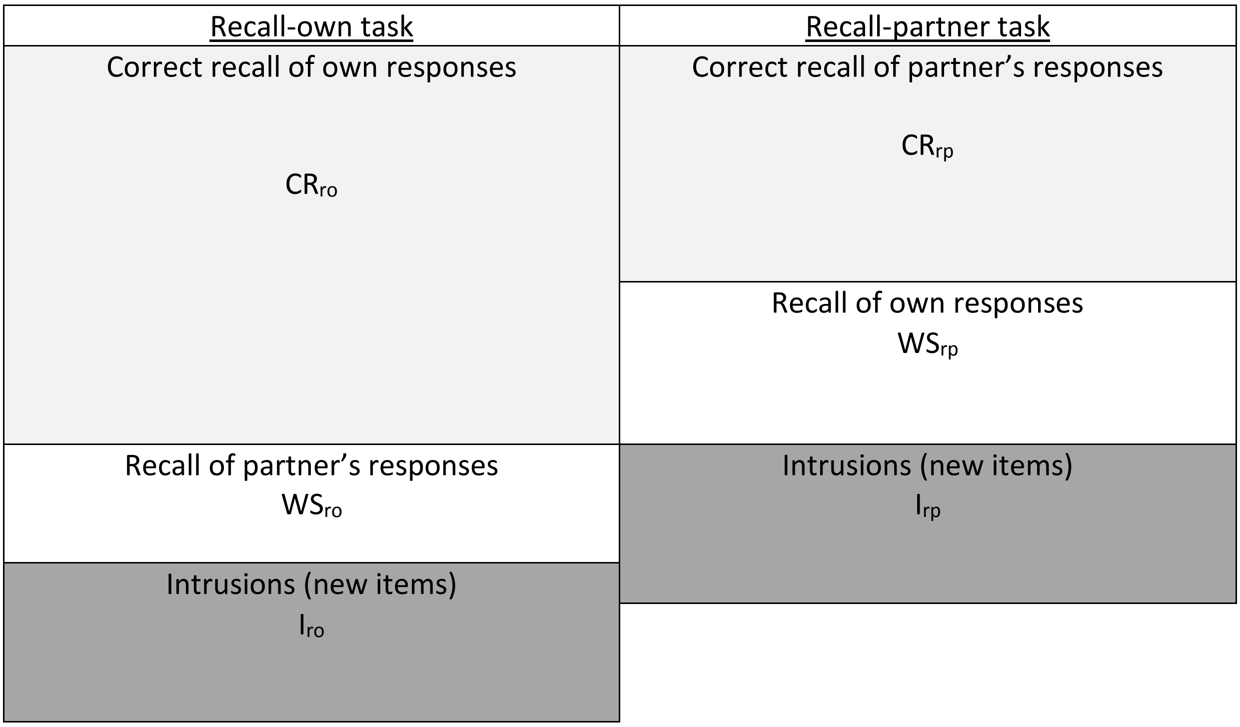
Table 3: Experiment 2b: The frequency of correct-source recall, wrong-source recall and intrusion errors written as own, or partner's responses in joint recall tasks across delay.

	1 day				1 week			
	Recall Own		Recall Partner		Recall Own		Recall Partner	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Correct Source	23.1	4.72	15.8	3.82	17.6	4.31	10.4	4.75
Wrong Source	1.53	1.12	1.42	1.08	3.15	2.50	3.63	2.33
Intrusions	1.42	2.29	2.68	2.87	3.42	3.55	5.21	3.69
% Source Error	6.1%	4.6%	7.5%	6.3%	12.1%	8.8%	19.4%	11.3%

Table 4: Experiment 3: The frequency of correct-source recall, wrong-source recall and intrusion errors in recall for recall-own and recall-partner tasks that were either generated (written anywhere on the response sheet), or included as task-compliant in the output.

	1 day				1 week			
	Recall-own		Recall-partner		Recall-own		Recall-partner	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Items included as task compliant								
Correct source	48.8	5.19	35.9	8.55	37.0	5.17	26.3	7.31
Wrong source	2.75	1.92	5.70	3.39	8.30	4.03	12.2	5.97
Intrusions	1.45	1.31	5.25	3.82	5.90	4.32	11.8	6.63
% Source Error	5.2%	3.6%	12.7%	8.4%	16.0%	6.9%	23.7%	7.9%
Items generated during retrieval								
Correct source	52.25	4.14	38.50	8.22	45.85	7.11	32.29	8.39
Wrong source	34.95	7.20	42.65	13.89	28.60	7.24	39.67	10.13
Intrusions	8.60	4.75	10.95	6.32	21.55	16.44	22.57	13.00
% Source Error	36.2 %	3.9%	45.2%	10.5%	30.2%	5.6%	42.6%	7.3%

Figure 1: Schematic representation of differing recall for recall-own (RO) and recall-partner (RP) tasks. The different areas in each column depicts the potential for different numbers of items to be correctly recalled (CR), recalled from the wrong source (WS), or to be produced as an intrusion (I) in each retrieval task.



Notes: These proportions of items can be used to derive the proportion of wrong-source errors, either as a function of the task, or as a function of the original source. For task-contingent analysis, the proportion of source errors is given by: Recall own = WS_{ro} / (CR_{ro} + WS_{ro} + I_{ro}), Recall partner = WS_{rp} / (CR_{rp} + WS_{rp} + I_{rp}). For source-contingent analysis, the proportion of source errors is given by: Own ideas = WS_{rp} / (CR_{ro} + WS_{rp}), Partner ideas = WS_{ro} / (CR_{rp} + WS_{ro}).

Figure 2: Experiment 3: The number of non-target responses generated during retrieval (Panel A), the proportion of these reported as task compliant (Panel B) and the resultant number of intrusion errors and wrong-source errors reported as task compliant (Panel C) for recall-own (RO) and recall-partner (RP) tasks tested after a delay of 1 day or 1 week. Error bars depict the standard error of the mean.

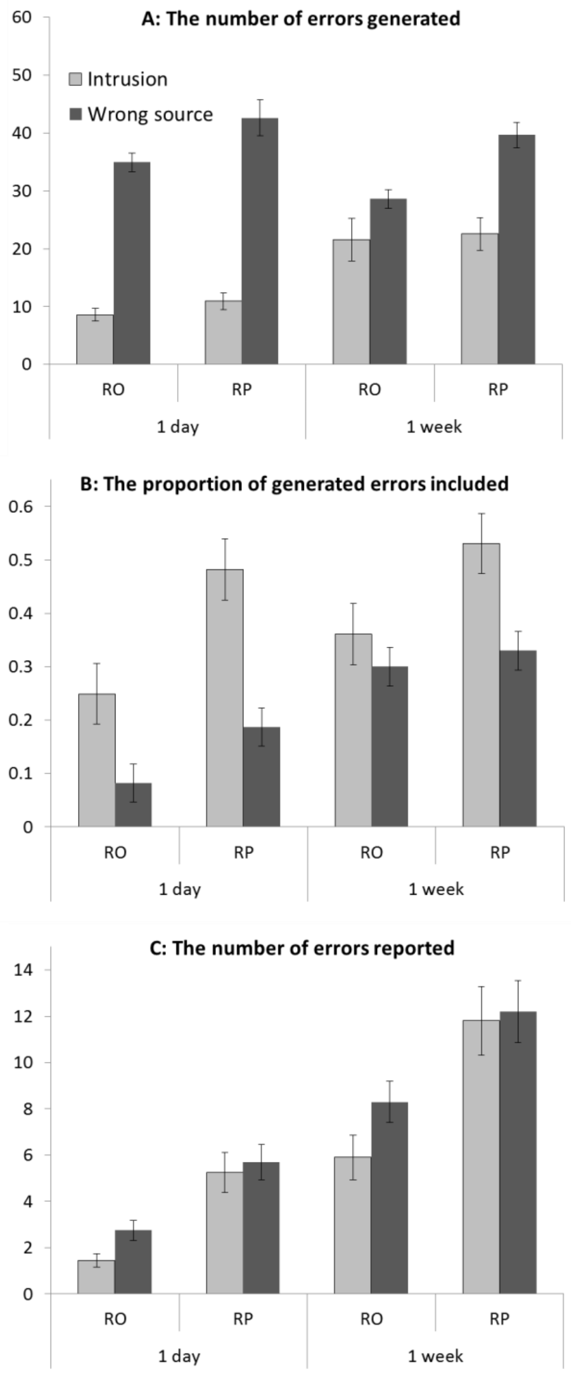
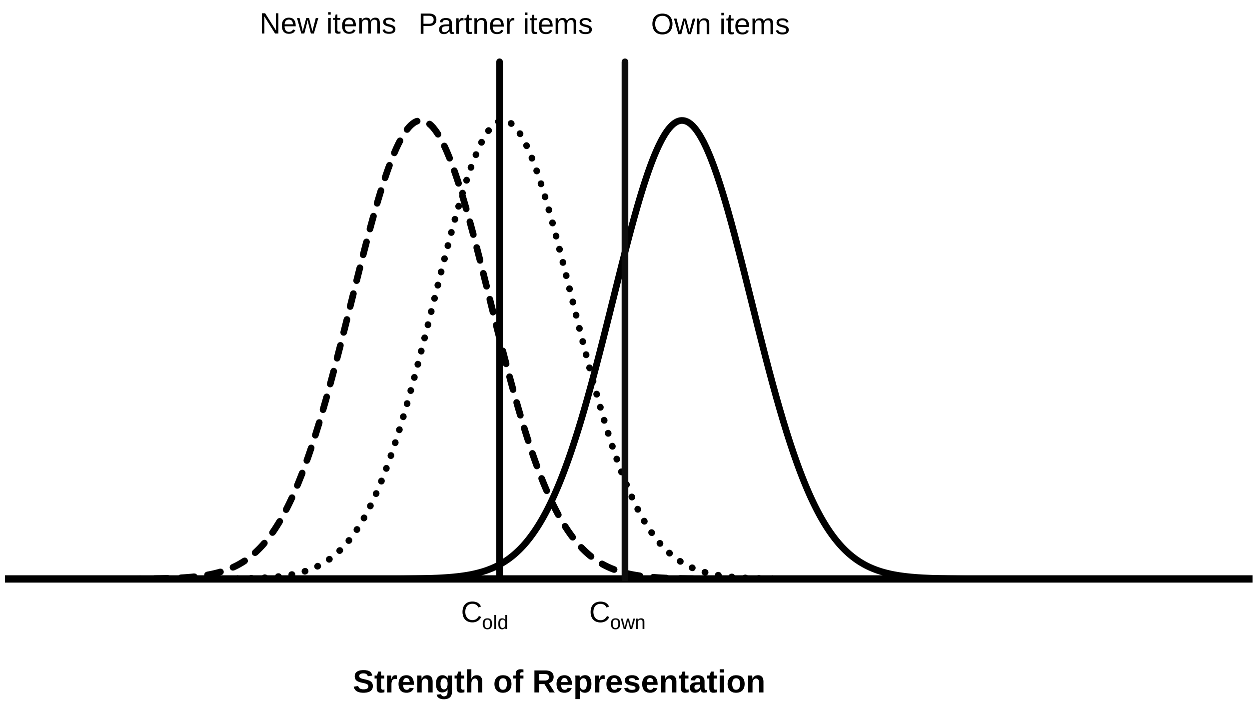


Figure 3: Schematic representation of the relative-strength model of unconscious plagiarism from Marsh and Landau (1995).



Notes: Hypothetical strength distributions for new ideas (dashed line), ideas generated by a partner (dotted line), and own ideas (solid line). The upper criterion (C_{own}) represents the strength threshold above which items are classified as own, and the lower threshold (C_{old}) represents the strength criterion at which items are judged as old. Items falling between the two criteria are judged to have come from a partner.