Psicológica (2015), 36, 17-36.

Exploring Retrieval Induced Forgetting with ad hoc categories

Patrícia Marinho and Josefa N. S. Pandeirada*

University of Aveiro, Portugal

The retrieval-practice paradigm has demonstrated that the act of selectively recovering some of the previously studied items from a category impairs the retrieval of the remaining items from that category, as compared to the retrieval of information from non-practiced categories (retrieval-induced forgetting); the practiced items are also better remembered than the items from non-practiced categories (a facilitation effect). This paradigm typically uses semantic categories, but its classic effects have been observed with other stimuli, such as lexical cues or ambiguous words. However, no study has tested this paradigm using ad hoc categories, a type of material that shares many characteristics with semantic categories; this was the goal of the present study. Our results replicated the facilitation effect, but we did not observe retrieval-induced forgetting. We discuss our results in light of the existing theories, suggest several factors that may underlie these results and propose future studies.

Forgetting is most typically considered a fault of memory rather than a useful characteristic. The critical importance and adaptive value of active forgetting mechanisms has been clearly expressed by Nairne and Pandeirada (2008) when discussing the frequent demand to "inhibit or suppress information in specific situations in which that information is not

* Acknowledgments: Both authors contributed equally to this work. This experiment was conducted to partially satisfy the requirements for Patricia Marinho's master dissertation. We would like to also thank the two anonymous reviewers and the editor of the manuscript for their suggestions that helped us improve this manuscript. Funding: Josefa N. S. Pandeirada was supported by the Portuguese Foundation for Science and Technology Fellowship (SFRH/BPD/81438/2011). This work was also partially supported by the Portuguese Foundation for Science and Technology (FCT) Grants PTDC/MHC-PCN/5274/2012 and IF/00058/2012 to Josefa Pandeirada. Corresponding author: Josefa N. S. Pandeirada. Department of Education, University of Aveiro. Campus Universitário de Santiago. 3810-193 Aveiro, Portugal. Telephone: +351 234 370 353. Fax: +351 234 370 640. Email: josefa@ua.pt

needed" (p. 189), and also the role it exerts in cognitive control; they were referring to a form of active forgetting: retrieval-induced forgetting (RIF). This forgetting mechanism is usually studied using the retrieval-practice paradigm, which has demonstrated that recalling an event or item triggers the forgetting of associated memories (Anderson, Bjork, & Bjork, 1994).

The retrieval-practice paradigm typically includes four phases. In the first phase participants study a list of category-exemplar pairs for a later memory test. In the second moment participants perform a cued-recall task for half of the exemplars from half of the categories presented in the study phase; this is the retrieval practice phase. During the retention period that follows, participants perform an unrelated distracting task. At the end, a final test occurs: in a cued-recall task participants are instructed to remember all the items presented in the experience. This procedure creates three types of items: practiced items - the items recovered in the retrieval practice phase (Rp+); competitive items - the items from the practiced categories that were not practiced in the retrieval practice phase (Rp-); and, baseline items - the items that belong to the unpracticed categories (Nrp). Furthermore, this procedure generates two effects in the final memory task: a facilitation effect and RIF. The first effect refers to a higher recovery of the Rp+ items compared to Nrp items, while RIF refers to a lower recovery of the RP- items compared to the baseline items (Nrp) (for a review see Anderson, 2003; Anderson & Levy, 2010). The phenomenon of highest interest in this paradigm is RIF as it demonstrates a specific form of forgetting.

Although RIF is mostly studied with semantically categorized lists of items, it has also been shown to occur with other types of materials. For example, it has been tested using sentences that shared a topic or relation among them (Anderson & Bell, 2001; Gómez-Ariza, Lechuga, & Pelegrina, 2005), using a combination of spatial and semantic cues (Bäuml, Zellner, & Vilimek, 2005), applying lexical common cues instead of semantic categories (Bajo, Gómez-Ariza, Fernandez, & Marful, 2006), and using ambiguous words (Shivde & Anderson, 2001). It has also been obtained with a large variety of final tests, such as category-cued recall tasks (e.g., Anderson et al., 1994), category-plus-stem recall tasks (e.g., Anderson, Bjork, & Bjork, 2000), and recognition tasks (e.g., Hicks & Starns, 2004). Finally, this paradigm has also been extended to various experimental contexts such as eyewitness and everyday events (e.g., Shaw, Bjork, & Handal, 1995). However, despite the robustness of its demonstration, the mechanisms that underlie RIF are still under discussion.

One of the theories of RIF asserts that an inhibition mechanism produces this form of forgetting. According to this inhibitory perspective, first introduced by Anderson et al. (1994) and Anderson & Spellman (1995), all the items that share a common cue (i.e., category) become activated by the cue during the retrieval practice phase. However, for the to-be-practiced items to be successfully recovered, an inhibitory process of the competing items comes into play diminishing their chances of recall in the final phase (Bäuml et al., 2005). The inhibition of the unpracticed items during the retrieval practice phase impairs the subsequent recall of these exemplars during the final recall (Storm, Bjork, Bjork, & Nestojko, 2006). The occurrence of inhibition depends of particular factors, such as response competition (Anderson & Levy, 2010; Storm & Levy, 2012) and the degree to which the memory trace tapped by the test matches the one inhibited during the retrieval practice phase (i.e., transfer-appropriate processing; Anderson, 2003). Several key findings have been claimed to support the inhibitory perspective. For example, Anderson and Spellman (1995) demonstrated that RIF occurs when the cues used in the final test differ from the ones used during retrieval practice phase (i.e., is cue-independent). Moreover, the effect occurs mostly when there is a selective recovery attempt during the retrieval practice phase (i.e., retrieval dependence); strengthening the exemplar-cue association by other means (e.g., repeated exposure such as extra study) is usually not enough despite the fact that the Rp+ items do benefit from it (strength independence; e.g., Anderson & Bell, 2001; Anderson, Bjork, et al., 2000). It has also been shown that RIF occurs irrespectively of whether the participant succeeds in recovering the correct item during the retrieval practice phase (e.g., Storm et al., 2006). Other studies have reported that RIF is moderated by the amount of competition caused by the Rp- items (competition or interference dependence; Anderson, Bjork, et al., 2000; Anderson et al., 1994).

Another theory that tries to explain this phenomenon is the associative theory. This theory argues that the association between the common cue (i.e., the category) and the practiced exemplars is strengthened during the retrieval practice phase. As a result, when attempting to recall the non-practiced items during the final test, in the face of the common cue, the practiced exemplars "intrude" persistently due to their faster activation, diminishing the likelihood of recalling the first. Recent studies based on the associative theory can explain some of the previous key findings that support the inhibitory perspective. For example, Verde (2013) has recently reported that the simplistic version of the associative theory, as described above, is unable to account for the retrieval dependence phenomenon but that a more sophisticated model can. In several experiments the author

manipulated the type of encoding during practice and showed this influenced the occurrence of RIF, a result predicted by his complex interference SAM-REM model. His results demonstrated that when the practice selectively enhances the encoding of features related to the categorical context of the information, the recall of the Rp+ is facilitated and interferes with the recall of the competitive items. However, when the practice selectively enhances the encoding of item features, recall of the Rp+ is also facilitated but there is no interference with the recall of the Rp-. These results indicate that the key for RIF to occur is not the retrieval dependence, but rather the fact that the typical retrieval-practice procedure enhances the encoding of categorical context features. Verde (2013) also suggested that inhibition does not readily explain other findings of RIF present in the literature (e.g., cue-independence).

Jonker, Seli, & MacLeod (2013) have proposed a different theory to explain RIF – the context-based account – arguing that the internal context also plays an important role in RIF. Specifically, they have showed that changing the context between the initial presentation of the items and their repeated study leads to RIF. On the other hand, if no context change occurs between the initial presentation of the items and their repeated study, or if the final task reinstates the initial learning context, RIF does not occur. These last results are at odds with the inhibitory account predictions.

Other studies have also indicated the existence of boundary conditions of RIF. For example, is has been demonstrated that RIF is moderated by the amount of integration that occurs at the time of encoding. This result has been associated with the existence of pre-experimental interconnections among the items that share a common retrieval cue (i.e., semantic integration) or to new relations revealed during the course of the experiment (i.e., episodic integration). For example, in the study by Anderson & McCulloch (1999), participants studied the category-exemplars under the standard encoding instructions, (i.e., they were instructed to study each pair relating the exemplar with its category label) or to rehearse each item with previously studied items, as well as to relate it to the category label (i.e., integrative-rehearsal condition). Across experiments the last condition reduced RIF, a result attributed to the integration among the items that occurred during encoding. Additionally, a questionnaire applied at the end of the experiment demonstrated that participants in the standard encoding condition spontaneously performed integration as well (what the authors called "self-reported integration"), which was even more likely to occur when more time was given during encoding.

More recently, Goodmon & Anderson (2011) have also demonstrated that one's prior knowledge about inter-relationships between

Rp+ and Rp- items prevents RIF from occurring and that the semantic integration can occur even when participants don't intentionally integrate the items. For example, in their fourth experiment, they used the material from Butler, Williams, Zacks, & Maki (2001) and demonstrated that their failure to obtain RIF was due to the introduction of preexisting associations beetween the practiced itens and their competitors (i.e., Rp- items) despite the fact that these associations were relatively weak. However, integration was protective of RIF only when it occurred between the competitor items and the target item. On the other hand, when associations were formed only among the Rp- items, integration did not protect from the competitors and, consequently, RIF emerged (Anderson, Green, & McCulloch, 2000; Goodmon & Anderson, 2011). Another contribution of the work by Goodmon & Anderson (2011) was the demonstration that semantic and episodic integration are two distinct moderators of RIF. Integration has been generalized to other stimulus with pratical implications, such as eyewitness events (Garcia-Bajos, Migueles, & Anderson, 2009).

As mentioned before, most studies of the retrieval-practice paradigm have used semantic categories. In the current study we used ad hoc categories: categories composed of semantically unrelated items, but that share a common theme. The exemplars of semantic categories have characteristics that co-occur naturally in the environment; for example, if we know that an exemplar of the category has feathers, we assume that it is much more likely to "fly" and "nest" than to "swim" or "possess gills." On the other hand, ad hoc categories derive from an objective or theme, so it is very unlikely that the items are naturally related or share common characteristics as occurs with semantic categories (Barsalou, 1983). Semantic categories are also well established in memory, have a stable presence in memory, and their taxonomic organization facilitates the retrieval of the corresponding items in memory tasks. In contrast, because ad hoc categories are formed only at the moment they are useful for a particular purpose they do not have stable representations in memory. This volatile presence in memory has implications on three levels: 1) the concept-to-instance associations are weak, that is, fewer ad hoc category items are recovered as compared to items from semantic categories; 2) weak instance-to-concept associations, meaning that participants only identify a set of exemplars as belonging to a certain category if the common theme is revealed; and, 3) weak associations among the concepts of the category, i.e., the features of ad hoc categories items are not clearly associated because they are rarely processed simultaneously. Consequently, the concepts of ad hoc categories show a lower accessibility when compared to the concepts of semantic categories (Barsalou, 1983).

Despite these differences, ad hoc categories comply with the minimum requirements for RIF to occur, such as the presence of a graded structure. This allows the selection of exemplars with a high frequency of production when people are prompted with the category name (e.g., "things that are green"); by using items that are at least moderately associated to the category we ensure exemplar competition. Importantly to the present study, the items from ad hoc categories share a common cue that can be used to recover the exemplars – the category name – which guarantees that a certain degree of competition should occur when the category label is presented for the retrieval of its exemplars (Barsalou, 1983). Free recall performance of ad hoc categories is also similar to that obtained with semantic categories when the category structure is provided (e.g., the category name) and participants engage in relational processing tasks (e.g., a category sorting task). Moreover, participants seem to rely significantly in the categorical structure of the information when outputting the information, in some cases to a similar extent than when using semantic categories; this is usually revealed by adjusted ratios of clustering, a measure that indicates the extent to which participants cluster their free recall according to a specific organization (ARC; e.g., Hunt & Einstein, 1981).

Providing that ad hoc categories share the properties of semantic categories that are necessary to obtain the typical effects produced by this paradigm, we expected to replicate them (i.e., the facilitation effect and RIF). Other aspects of the procedure required for the effects to occur were also considered. First, participants were specifically asked to memorize the items by trying to relate each item with its category, thus making the category label a common cue and providing an organizational structure to the material (see Anderson et al., 1994). We also provided one more second than what is typically used during study and recovery of the exemplars, allowing extra time for the participant to establish a stronger relation between the item and its ad hoc category (e.g., Anderson, Bjork, et al., 2000); this is a procedural change frequently applied when materials other than semantic categories are used in the retrieval-practice paradigm (see, Anderson & Bell, 2001; Ciranni & Shimamura, 1999; Gómez-Ariza et al., 2005; Shivde & Anderson, 2001). Second, to assure the occurrence of competition during the retrieval practice phase, the following aspects were considered: 1) the average typicality of the items in each category was at least moderate; 2) the retrieval cue appealed to the relation between the category label and its exemplars; 3) care was taken to prevent the usage of items that would be highly related among each other within a category (e.g., Anderson et al., 1994; Goodmon & Anderson, 2011); and, 4) the selective recovery occurred three times for each item increasing the association

between the item and its category, a frequent procedural detail that is also maintained when other materials/contexts were applied (see, Anderson & Bell, 2001; Ciranni & Shimamura, 1999; Gómez-Ariza et al., 2005; Shaw et al., 1995; although see MacLeod & Macrae, 2001). We also used a memory task that robustly produces the effects both during the retrieval practice and the final memory task – the category-plus-stem recall task (for a review, see Anderson, 2003).

METHOD

Participants. Thirty psychology students from the University of Aveiro (27 women) aged 18-41 years (M = 21.8, SD = 4.67) participated in the experiment voluntarily and, in some cases, in exchange for credits in Curricular Units. This sample size is similar (and even larger in some cases) to that used in other studies where RIF was successfully obtained both using the typical paradigm (e.g., Bäuml et al., 2005; N = 24) as well as variations of the paradigm that used different stimulus (e.g., Bajo et al., 2006; N = 20 and N = 26, in Experiments 2 and 3, respectively; Gómez-Ariza et al., 2005; N = 32 in Experiment 1). All participants agreed to an informed consent form before beginning the experiment. They were tested individually in sessions with a maximum of 12 participants that lasted approximately 40 minutes each.

Materials. Eight ad hoc categories were selected from the Portuguese norms for ad hoc categories (Pandeirada & Marinho, Under preparation; see Appendix 1 for procedural details of this work): "things women wear", "things people put on walls", "things that make noise", "things made mostly of plastic", "things people keep in their pockets", "things that have smell", "things that are flammable", and "things dogs chase". The first six categories were the experimental categories and the last two were buffer categories (these were not considered in the results). Categories were selected such that they would constitute distinct and non associated categories (i.e., each category corresponded to a different theme and the corresponding exemplars had very low probability of being considered members of another category). The categories were labeled by a short sentence that would clearly identify it (e.g., "things women wear").

For each category 6 exemplars were selected according to various constraints. The first was that exemplars were unambiguous, (i.e., they would very hardly be considered to belong to another of the used categories), consisted of a single word, had similar lengths, and the retrieval

cue of each exemplar of a given category (i.e., their first two letters) was unique within that category (c.f., Appendix 2). Also, on average, exemplars were moderately frequent in their categories (average frequency = 26.7% and average rank order = 9). Both the average frequency and the rank order of the items have been used in the literature as indicators of the typicality of the items. Making the typicality of the item at least moderate ensures there will be competition among the items during the selective recovery phase (Bäuml, 1998; Bäuml & Hartinger, 2002). Finally, to prevent the usage of inter-exemplar retrieval strategies during the testing phase, exemplars with strong item-to-item associations were avoided (e.g., exclusion of the items "cats" and "(other) dogs" for the category "things that dogs chase").

Design. The status of the exemplars during the testing phase was manipulated within-subject: the Rp+ corresponded to half of the items from 4 practiced categories; the Rp- exemplars corresponded to the other half of the items from the practiced categories; and, the Nrp exemplars belonged to the remaining two non-practiced categories. The status of the categories (i.e., practiced *vs.* not practiced) and the statuses of the exemplars from the practiced categories (i.e., Rp+ and Rp-) were counterbalanced across participants so that all the categories and all the items participated in each condition. Additionally, for each of the counterbalancing versions, five different orders of presentation of the categories during the final testing phase were created; the first to-be-retrieved category was always a buffer category to allow familiarization with the task.

Procedure. The procedure was applied using the E-Prime program (Schneider, Eschman, & Zuccolotto, 2002). On arrival at the laboratory, participants were randomly assigned to one of the versions of the experiment. The experiment was conducted in four phases as described next.

Study phase. In this phase, participants were informed that (a) they would be participating in a memory and reasoning task; (b) they would be individually presented a category label along with an exemplar of that category; and, (c) they would have six seconds to study each category-exemplar pair, and should spend this time trying to relate the item to the respective category (a 1 sec ITI was used between items). Aside from the timing information in our instructions, the remaining instructions mimicked the ones used by Anderson et al. (1994) where RIF was obtained. The category-exemplar pairs were presented in lowercase on the center of the computer screen, with the category label appearing above the category

exemplar. The order of presentation of the category-exemplar pairs was randomly pre-determined and the same for all participants, except that at the beginning and end of the study phase two items from each of the buffer categories were presented, and no two items from a given category were presented adjacently.

Retrieval practice phase. Upon completion of the study phase, participants were instructed to perform the retrieval practice phase. In this phase, participants were told that the name of one of the previously-studied categories would be presented on the screen along with the first two letters of one of its exemplars also previously studied. They were also told that their task was to complete the two letters with one of the items from that category that they had previously studied and that they would have 8 seconds to recover each item and type it using the keyboard (a 1 sec ITI was used between items). The display was similar to the one used in the study phase, except that now only the two initial letters of the exemplar were revealed and followed by an underline of constant size to prevent any additional cue that could indicate the length of the to-be-retrieved word. Thus, a category-plus-stem cued task was used, where the name of the category and the two initial letters of the to-be recovered item were used as retrieval cues (e.g., "things women wear" so ____). Fifteen of the previously learned items (three exemplars of each of the to-be-retrieved categories plus three exemplars of one buffer category) were presented in this phase. Order of presentation of the items was randomly pre-determined except that at the beginning and end of the list an item from the buffer category was presented and no two items from the same category were presented adjacently. This form of presentation was repeated three times.

Retention period. The retrieval practice phase was followed by a 5 minute retention period during which participants had to discriminate whether a number presented in the center of the screen was odd or even. Responses were provided by pressing the key "I" for an odd number and the "P" key for an even number. The numbers ranged between 0 and 9 and their order of presentation was random.

Testing phase. In this phase, participants were given the same instructions as in the retrieval practice phase. The form of presentation of the category-plus-stem, the timing to respond and interval between stimuli were also the same. However, the recovery of the exemplars was now blocked: participants were asked to recover all the items from a given category and, only then, moved to another category. The order of presentation of the items within each category was randomly determined for each participant following the procedure used in previous studies were RIF was obtained (e.g., Butler et al., 2001).

RESULTS

The significance level for all statistical comparisons was set at $p < .05$.

Table 1 shows the percentage of correct responses for each type of item during the final test. A repeated measures ANOVA was conducted to analyze the facilitation effect (Rp+ > Nrp). The result revealed a significant difference confirming that the practiced items were better recalled than the baseline items, $F(1,29) = 6.721$, $MSE = .014$, $p = .015$, $\eta_p^2 = .19$. Using the same analysis, the comparison for the final recall of the competing items (Rp-) and baseline items (Nrp), did not result in a significant difference; $F(1,29) = 0.037$, $MSE = .012$, $p = .85$, $\eta_p^2 = .001$, not replicating the typical RIF. Although we adopted measures to ensure competition would be present among the items of a given category, we further explored our data considering that the categories with items of higher average frequency would afford more competition than categories with lower frequency items. The relevant data are described also in Table 1 where RIF was calculated for each specific category. As can be seen, in 4 out of the 6 categories, Rp- items were actually better recalled than the Nrp items, the opposite of RIF. A pattern of results consistent with RIF was obtained in the remaining two categories; one case refers to a category with items of high average frequency, but the other is actually the category with the lowest level of item frequency. Therefore, RIF does not seem to be related to the typicality of the items within a given category, at least when ad hoc categories are used.

One can also wonder if our experiment was powerful enough to detect RIF. A brief summary of RIF reported in several papers using both semantic categories, as well as other types of material, indicates an average effect size of Cohen's $d = 0.68$[1]. A post-hoc power analysis revealed that the likelihood of failing to obtain an effect of this size with a sample of 30 participants is virtually zero, indicating that our null result is not likely due to a lack of power to detect it.

One aspect that seems to influence the likelihood of recall during the final test is the performance during the retrieval practice phase. Storm et al. (2006) demonstrated that RIF occurs despite the successful recovery in the retrieval practice phase. However, they also reported data suggesting a trend for higher RIF when the performance during the retrieval practice was

[1] Cohen's d effect sizes for the data from other studies were calculated based on the reported averages and standard deviations. These were calculated for studies that provided the averages and SDs that allowed us to make this calculation.

Table 1. Average percentage of correct responses (and *SDs*) during the testing phase, for the overall material and for each category.

	ACF	Rp+	Nrp	Rp-	RIF
Overall data	26.7	82.5 (13.0)	74.4 (14.0)	75.0 (15.0)	-0.6
Data by category					
Things people keep in their pockets	38.3	86.7 (22.7)	78.3 (13.7)	85.0 (17.0)	-6.7
Things women wear	35.5	91.2 (18.7)	86.7 (10.5)	76.7 (21.9)	10.0
Things that have smell	24.8	70.0 (28.4)	60.0 (25.1)	66.7 (30.6)	-6.7
Things people put on walls	23.9	89.5 (15.9)	75.0 (23.9)	76.7 (28.8)	-1.7
Things that make noise	22.5	81.7 (20.1)	75.0 (16.2)	80.0 (22.7)	-5.0
Things made mostly of plastic	15.0	75.0 (23.9)	71.7 (13.7)	65.0 (25.3)	6.7

Notes: ACF: Average category frequency. Rp+: exemplars of practiced categories during the retrieval practice phase. Rp-: exemplars of practiced categories not recovered during the retrieval practice phase. Nrp: exemplars from unpracticed categories. RIF: difference between the mean recovery of Nrp and Rp- exemplars.

lower (although the effect did not reach statistical significance). To explore this phenomenon we further analyzed our data by looking at the participants' performance in the final test as a function of their performance during the retrieval practice. The average success rate of recovery in this phase was 81% (*SD* = 12.0%), which is similar to that obtained in other studies (e.g., Anderson, Bjork, et al., 2000; Anderson et al., 1994; Butler et al., 2001). We then performed a median split based on this average performance, creating the group of high generators (*M* = 90%, *SD* = 7.0%) and the group of low generators (*M* = 72%, *SD* = 7.0%). Descriptive data for each group and type of item are presented in Table 2. Mixed analysis of variance with generation group as a between-subjects variable on the final recall performance was conducted to explore the effect of the retrieval practice performance on RIF. The results revealed no significant main effect of type of item, $F(1,28) < 1$, of group, $F(1,28) = 3.65$, $MSE = .029$, $p = .066$, $\eta_p^2 = .115$, nor interaction $F(1,28) = 3.25$, $MSE = .012$, $p = .08$, $\eta_p^2 = .104$. Although the interaction did not reach the statistical significance, our

descriptive data go in the same direction as those reported by Storm et al. (2006): the low generators manifested a higher RIF than the high generators. In our study, the data from the high generators actually go in the opposite direction of RIF. When analyzing the condition where the generation task was impossible, participants from their study who generated fewer responses were also more prone to a higher RIF.

Table 2. Average percentage of correct responses (and *SDs*) during the testing phase by group based on performance during retrieval practice (high and low generators).

	Rp+	Nrp	Rp-	RIF
High Generators	91.7 (8.0)	76.1 (11.0)	81.7 (11.0)	-5.6
Low Generators	73.3 (11.0)	72.8 (17.0)	68.3 (16.0)	4.5

Notes: Rp+: exemplars of practiced categories during the retrieval practice phase; Rp-: exemplars of practiced categories not recovered during the retrieval practice phase; Nrp: exemplars from unpracticed categories; RIF: difference between the mean recovery of Nrp and Rp- exemplars.

Storm et al. (2006) also suggested that the participants who generated fewer items might have put a higher effort in the retrieval search process which, in turn, increased the suppression of the competitor items. If this is the case, one might expect these participants to take longer to retrieve their responses. Indeed, when measuring how long participants took to start generating a response during the retrieval practice phase, participants with the lowest level of performance took longer to initiate their responses (M = 3705.2, SD = 608.7), as compared to the ones with better performance (M = 3449.2, SD = 409.2); however, this difference did not reach a significant level, $t(28) = 1.35$, $p = .19$.

DISCUSSION

In this study we implemented the retrieval-practice paradigm, but using a different material than the typically used semantic categories: we used ad hoc categories. Two classical results are usually obtained from this paradigm: the facilitation effect and RIF. The present study replicated the facilitation effect using ad hoc categories: the items retrieved during the retrieval practice phase (Rp+) were better recalled in the final test than the items from categories never practiced (Nrp). On the other hand, we did not obtain RIF, that is, better recovery of the items from never practiced categories (Nrp), as compared to the items from practiced categories but that were not retrieved during the retrieval practice phase (Rp-). One possible reason for the absence of this effect is that the Rp- items were not inhibited during the retrieval practice phase, which could be related to the nature of the categories used. As mentioned in the introduction, ad hoc categories do not respect the correlational structure of the environment and have an unstable presence in memory, as opposed to what is characteristic of semantic categories (Barsalou, 1983). According to the inhibitory account of RIF, this effect depends on the occurrence of competition during the selective retrieval, and this competition is closely related to how the exemplars are related to the common cue (i.e., the category name or label; Anderson, 2003). Although we controlled the level of association of the items to their respective categories by selecting items moderately associated to the categories, we can wonder whether participants identified all the exemplars as belonging to the respective category. If the category was not considered a good retrieval cue to some of its exemplars, a lower level of competition could occur during the retrieval practice phase and, consequently, the level of competition among the items of that same category would be reduced. However, it has been shown that participants, in the presence of a set of items and their respective category (in this case ad hoc categories), have a high level of agreement as to how good the item is for its category (subjective typicality; Barsalou, 1983). Therefore, having chosen moderately high frequent exemplars of each category, and knowing that there is a high agreement on subjective typicality in ad hoc categories, it is unlikely that the participants had difficulty relating the presented exemplars with their respective categories.

Another element that suggests that the retrieval cue (category label-plus-word stem) was effective in cueing the corresponding exemplars is the highly successful performance in the retrieval practice phase. In fact, this level of performance was similar to the one obtained in studies that have used semantic categories (e.g., Anderson, Bjork, et al., 2000; Anderson et

al., 1994; Butler et al., 2001). Still, because we have some variation in the average typicality of our categories, we looked at RIF for each category. Our results revealed no clear relation between the average typicality of the category and the occurrence of RIF. Interestingly, for the category with the higher typicality average where the pattern of results is consistent with RIF – the category "things women wear" – we can find most of its items in the semantic category of "articles of clothing" from the Van Overschelde, Rawson, & Dunlosky (2004) category norms; in fact, 4 of our items belong to this category and have an average frequency of production of 0.39. So, considering this category is quite similar to the semantic categories used in the typical RIF studies, this result suggests our procedure likely is capable of producing RIF. At the same time, it suggests that there is something related to the nature of ad hoc categories that is preventing the effect from occurring.

The usage of semantically categorized lists in the typical retrieval-practice paradigm studies encourages the similarity processing, often through categorical relationship, which increases the probability of competition during the retrieval practice phase. This processing is not optimal to memory because it introduces difficulties in differentiating among the particular items of a given category at retrieval. Response competition is triggered in response to this difficulty allowing the target items to be retrieved, thus producing RIF (Smith & Hunt, 2000). We question whether the ad hoc categories are capable of inducing this type of processing to the same extent as semantic categories. As reported by some authors, these categories are characterized by producing a lower processing advantage comparatively to the semantic categories (Barsalou, 1983). Thus, it is possible that the amount of similarity processing induced by the ad hoc categories during the selective recovery was not sufficient to produce RIF. On the other hand, when the encoding task focuses on the categorical organization, their recall can actually exceed that of semantic categories that underwent a similar encoding task (e.g., Hunt & Einstein, 1981). The authors suggested this advantage might accrue from the combination of item-specific (afforded naturally by the "unrelated" nature of the ad hoc categories) and relational processing (afforded by the relational encoding task). Item-specific processing is also sometimes related to a form of distinctive processing, which is known to mitigate RIF due to the reduction of response competition (Smith & Hunt, 2000). Furthermore, research has shown that participants are able to take advantage of the categorical structure of this type of material to a similar extent that they do with semantic categories (e.g., Hunt & Einstein, 1981).

One other procedural detail implemented in many studies where RIF was obtained is the manipulation of the typicality of the Rp+ and Rp- items (e.g., Anderson et al., 1994; Garcia-Bajos et al., 2009). Specifically, the procedure that most likely leads to RIF includes the usage of high frequency items for RP- items and low frequency items for Rp+ items. We did not control for this element in our study but rather implemented a counterbalancing procedure that guaranteed every item participated equally as an Rp+ and Rp- item controlling for any item-specific characteristics that could affect their memorability (e.g., concreteness, familiarity, etc). However, future studies should implement this manipulation in order to clarify its role in RIF with this specific type of material.

In our study we allowed 6 seconds to study each category-exemplar pair, 1 second longer than what is typically used. This increase in study time has been applied in studies that obtained RIF using other types of materials (see Anderson & Bell, 2001; Gómez-Ariza et al., 2005) making it unlikely that this change in our procedure was responsible for the present absence of RIF. On the other hand, this extra time also provides more opportunities for participants to spontaneously find associations among the items of a given category (i.e., integrate information), which can prevent RIF from occurring as reviewed in the introduction. Indeed, similar levels of performance for the Rp- and the baseline items when integration occurs between the Rp+ and the Rp- have been reported (e.g., Anderson & McCulloch, 1999). It is possible that this extra second in learning time also promoted spontaneous integration in our study preventing RIF from occurring.

Goodmon & Anderson (2011) also found that uncontrolled preexisting associations between practiced and nonpracticed items, even when weak, are sufficient to produce semantic integration (see their Experiment 4) and, consequently, RIF does not emerge. As reported by Anderson et al. (1994), semantic associations must be intentionally controlled to prevent the action of integration. However, because ad hoc categories include items that do not a priori share characteristics (as occurs with semantic categories; Barsalou, 1983), they are not as prone to strong item-to-item associations. Furthermore, the items from the different categories were intermixed during study, which makes it even harder to establish these associations.

Barsalou (1983) points out that ad hoc categories can lose their status when they are processed frequently as their category concepts, concept-to-instance associations, and instance-to-concept associations become more stable in memory, thus becoming more similar with semantic categories. The loss of the ad hoc status in categories can make semantic integration more prone to occur. It seems very unlikely that a single presentation of the

ad hoc items in the intermixed procedure would cause such an effect. Still, because of the ad hoc nature of the categories, participants might have tried to find communalities among the items of each category during the initial presentation in an effort to "create" each ad hoc category as the corresponding items were being presented. This could be a form of spontaneous integration that would inhibit RIF. This is a question that could be addressed in a post-experience questionnaire in future research.

Another finding that can be related to the integration hypothesis is the data on the low *vs.* high generators during the retrieval practice phase, where a pattern of results consistent to RIF was obtained for the first but not for the second group. The low generators (i.e., participants with lower performance during the retrieval practice phase) may have had more trouble relating the items with its respective category as well as with the remaining items of the category. The opposite might have happened with the high generators who could have engaged in some level of integration of the items with its respective category. Remember that both semantic integration and episodic integration seem to protect against RIF (Anderson & McCulloch, 1999). In the current study, by virtue of the nature of the material, it is possible that participants engaged in some form of episodic integration by trying to make sense of the relation between each item and a given category that did not pre-existed in their memory. At this point, our data are insufficient to completely explore this hypothesis; as proposed before, a post-experiment questionnaire in future research could help shed light into this question.

On the other hand, these results also speak to the potential role the level of performance during the retrieval practice phase might have in obtaining RIF. Our data indicated that RIF is more likely to occur for the low generators, which is consistent with the data reported by Storm et al. (2006). The response-time data are also in agreement with their suggestion of a higher effort to retrieve the items by these participants which, in turn, might increase the level of competition among the items. These aspects should deserve more attention in future studies as they might help understand this phenomenon. Of note, our current high *vs.* low generators analysis should be regarded only as exploratory given the low sample sizes involved.

In short, the facilitation effect usually observed in the retrieval-practice paradigm with semantic categories was replicated using ad hoc categories. On the other hand, we failed to replicate RIF, the most important outcome of this paradigm. To the best of our knowledge, no previous studies have used ad hoc categories, and our study provides a first exploratory look into this question. The existing research on this paradigm

does not allow a firm conclusion about the underlying causes of our failure to obtain RIF. The most plausible reasons relate to the nature of the material used and its impact in the processes usually considered to promote and/or inhibit RIF from occurring, such as competition among the items or integration. Our study identified what seems to be a boundary condition for RIF but it raises more questions than answers because, at this point, we are unable to clearly identify specific causes for our results. Our results should be seen as exploratory and only replication using other ad hoc categories will help establish this conclusion and improve our knowledge about RIF.

REFERENCES

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*(4), 415-445. doi: 10.1016/j.jml.2003.08.006

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130*(3), 544-570. doi: 10.1037/0096-3445.130.3.544

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7*(3), 522-530. doi: 10.3758/BF03214366

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20*(5), 1063-1087. doi: 10.1037/0278-7393.20.5.1063

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology. Learning, Memory & Cognition, 26*(5), 1141-1159. doi: 10.1037//0278-7393.26.5.1141

Anderson, M. C., & Levy, B. J. (2010). On the relationship between interference and inhibition in cognition. In A. Benjamin, B. (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 107-132). New York: Psychology Press.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory & Cognition, 25*(3), 608-629. doi: 10.1037/0278-7393.25.3.608

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model. *Psychological Review, 102*(1), 68-100. doi: 10.1037/0033-295X.102.1.68

Bajo, M. T., Gómez-Ariza, C. J., Fernandez, A., & Marful, A. (2006). Retrieval-induced forgetting in perceptually driven memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1185-1194. doi: 10.1037/0278-7393.32.5.1185

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*(3), 211-227. doi: 10.3758/BF03196968

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *Journal of Experimental Psychology, 80*(3 Pt.2), 1-46. doi: 10.1037/h0027577

Bäuml, K.-H. (1998). Strong items get supressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review, 5*(3), 459-463. doi: 10.3758/BF03208822

Bäuml, K.-H., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10*(3), 215-224. doi: 10.1080/09658210143000362

Bäuml, K.-H., Zellner, M., & Vilimek, R. (2005). When remembering causes forgetting: Retrieval-induced forgetting as recovery failure. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31*(6), 1221-1234. doi: 10.1037/0278-7393.31.6.1221

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory & Cognition, 27*(5), 1314-1319. doi: 10.1037//0278-7393.27.5.1314

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1403-1414. doi: 10.1037/0278-7393.25.6.1403

Garcia-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory, 17*(1), 92-103. doi: 10.1080/0958210802572454

Gómez-Ariza, C. J., Lechuga, M. T., & Pelegrina, S. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33*(8), 1431-1441. doi: 10.3758/BF03193376

Goodmon, L. B., & Anderson, M. C. (2011). Semantic integration as a boundary condition on inhibitory processes in episodic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 416-436. doi: 10.1037/a0021963

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11*(1), 125-130. doi: 10.3758/BF03206471

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 497-514. doi: http://dx.doi.org/10.1016/S0022-5371(81)90138-9

Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review, 130*(4), 852-872. doi: 10.1037/a0034246

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12*(2), 148-152. doi: 10.1111/1467-9280.00325

Nairne, J. S., & Pandeirada, J. N. S. (2008). Forgetting. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 179-194). Oxford: Elsevier.

Pandeirada, J. N. S., & Marinho, P. O. (Under preparation). *Portuguese norms for ad hoc categories*.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E- Prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*(2), 249-253. doi: 10.3758/BF03210965

Shivde, G., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 175-190). Washington, DC: American Psychological Association.

Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition, 28*(4), 503-508. doi: 10.3758/BF03201240

Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review, 13*(6), 1023-1027. doi: 10.3758/BF03213919

Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition, 40*(6), 827-843. doi: 10.3758/s13421-012-0211-7

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language, 50*(3), 289–335. doi: 10.1016/j.jml.2003.10.003

Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory & Cognition, 39*(5), 1433-1448. doi: 10.1037/a0032975

APPENDIX 1

Short description of the procedure used in the normative study.

We began by creating a list of potential ad hoc categories based on previous studies that have used ad hoc categories (e.g., Barsalou, 1983; Hunt & Einstein 1981). The fourteen categories identified as most commonly used through this method were divided into two sets of seven categories in order to prevent overlap responses between categories. An online questionnaire was constructed to collect data for each of these sets. Additionally, three different presentation orders for the categories in each questionnaire were created to minimize interference in responses across categories. These questionnaires were made available online via the e-learning platform of University of Aveiro and were advertised in several universities in Portugal by email. The version received by the participant was randomly determined.

The initial instructions informed participants that the goal of the questionnaire was to collect information on items or objects that people commonly consider to belong to multiple categories or classes. After participants consented to their participation and supplied several demographic data (e.g., age and sex), specific instructions for generation of the exemplars were provided. These instructions were based in the Battig & Montague (1969) normative study but were adapted to the ad hoc categories. The generation of the exemplars for each category was

intercalated with three questions regarding the process and strategies of generating exemplars for the just presented category (e.g., these questions were repeated after generating the items for each of the seven categories). This information was not considered in the selection of the exemplars for the current study.

Valid data from one hundred and ninety-two Portuguese-speaking subjects (M_{age} = 35.02 years, SD = 3.44) were collected in this study. Responses for each category were analyzed considering the frequency of production of each exemplar as well as the order in which the item was produced.

APPENDIX 2

List of categories and respective exemplars used in the study, along with the average frequencies of generation and frequencies of exemplares.

	Things women wear	Things people put on walls	Things that make noise	Things made mostly of plastic	Things people keep in their pockets	Things that have smell
Average Frequencies	35.5%	23.9%	22.5%	15.0%	38.3%	24.8%
Exemplars and respective frequencies	bra (56.4%)	picture (94.7%)	television (47.0%)	bags (38.6 %)	wallet (64.3%)	flowers (66.7%)
	socks (52.1%)	mirror (28.0%)	radio (31.8%)	packing (23.7%)	coins (53.2%)	fruit (24.4%)
	shorts (37.0%)	hangers (15.9%)	honk (31.3%)	boxes (13.5%)	money (45.0%)	shampoo (18.3%)
	tops (37.0%)	nails (14.8%)	motorcycles (25.8%)	bucket (13.0%)	bills (24.0%)	trash (18.0%)
	bikinis (19.4%)	frames (14.3%)	music (16.7%)	bowls (4.3%)	documents (17.0%)	scented soap (13.3%)
	pajamas (9.1%)	decorations (9.5%)	bell (11.6%)	filers (0.5%)	hard candy (16.4%)	deodorant (8.9%)

(Manuscript received: 10 March 2014; accepted: 11 September 2014)