ORIOLE MANUSCRIPT REVIEW HISTORY
REVIEWS (ROUND 1)

Editor Decision Letter

Thank you for submitting your manuscript to JCR. Your manuscript was read by three experts who provided constructive and complementary insights into your research. An equally expert Associate Editor integrated their comments and provided a set of additional comments. This letter provides my decision on the manuscript based on the review team’s guidance. A Trainee Reviewer also commented on your paper. Although that reviewer’s comments did not figure into my overall decision regarding the manuscript, that reviewer’s comments were constructive and helpful.

The review team thinks you have an interesting set of results and that you add a novel theoretical perspective (accountability) to bear on why experts may have more recall errors than novices. Your paper was well written and was a pleasure to read. Yet at present, neither the conceptual story nor the results are as clear as we would like. Conceptually, the alignability story is complex and needs more clearly explained in your theory section (Reviewer C, Trainee Reviewer). The accountability explanation needs greater prominence and integration in your theorizing (Reviewer C, Trainee Reviewer). Indeed, Reviewer C asks whether the alignability argument itself is needed. Could you instead provide a simpler story by suggesting that experts make more recall errors (and also have more recall successes) because they feel more accountable? In other words, is it alignability or accountability the critical issue in your paper? The answer to this question may depend on the extent to which you can respond to the AE’s concern that you have “overstated the novelty of your contribution by ignoring prior marketing research showing the expertise, involvement and accountability can increase bias”.

If the type of recall error (and hence alignability) really matters, then you will need to (a) clarify their differences earlier in the paper (Reviewer A, Trainee Reviewer) and make some changes with respect to your discussion of (b) the types of recall errors (see the AE), (c) the distinction between them (Reviewer A) and (d) the coding scheme used to characterize them (see Reviewer B). Reviewer B provides extremely useful direction on this latter issue. That reviewer’s comments coincide with that of the AE who suggest that the data needs to be recoded and the analyses need to be redone.

The reviewers also raise additional issues and suggest that you provide greater justification to the use of subjective vs. objective expertise and the use of free recall vs. cued recall or recognition. Better reporting of the results is desired (see Reviewers B and C). Sobel tests should be added to test mediation effects, and study 5 needs better justification. Study 1 can probably be dropped
(see the AE and Reviewer A), however, if you can justify its retention we might feel differently. Reviewers A and the Trainee reviewer ask whether you can determine whether these recall errors stem from encoding and retrieval. To the extent that you have data to support this issue its inclusion would be helpful. However, it is not the primary focus of the paper. You note this issue in your discussion section. Perhaps you can share some tentative hypotheses about ideas you may have on this issue in the Discussion section. Each of the issues noted in this paragraph are quite doable.

The reviewers were quite divided on your paper, with opinions ranging from conditionally accept to reject and resubmit. The AE recommends a revision which I support. However, I must state that this revision is very risky and I can make no claims about how it will fare in the next round of reviews. The positioning and essential contribution of the paper needs to be clarified, and the data may need to be recoded and reanalyzed. These are significant issues. Absent seeing your next submission it is impossible to predict how the next submission will fare. Still, I believe your paper has potential (and the review team agrees). As such, I hope that you will take on the challenge of addressing the significant issues that remain and submit a substantially revised paper.

If you do decide to revise the paper for JCR (and I hope you do), I plan to send the manuscript to the same review team. They like your work and have provided excellent comments. Hence, I think you will be advantaged by having them (vs. a new set of reviewers) see your next submission. As such, please submit a set of comments to the AE that describe how your revision addresses the comments raised by the team. If you wish, you may also submit an additional set of comments to each of the individual reviewers that indicates how you have responded to their specific comments. Although there is page limit to such comments, we have found that short and to the point responses (often less than 3 pages per reviewer) often suffice.

I hope these comments are clear. If not, please let me know. Thank you for submitting your work to JCR. Best wishes in your future work on this topic.

AE Report

Comments to the Authors:

This manuscript reports the results of five experiments showing that experts (vs. novices) are more likely to generate false memories about products due to schema-based inferences and due to accountability-driven impression management related concerns. Three knowledgeable reviewers found the studies interesting, however, all three expressed concerns about the coding scheme that was used to categorize different types of false memories. I share this concern and believe that a more appropriate coding scheme is needed. Furthermore, Reviewer C argues that the results can be explained more parsimoniously by the simple hypothesis that experts remember more.

Many of the reviewers’ concerns can be reduced by rethinking the false memory coding scheme. The type of false memories examined in this research are different from the type of false memories examined by Roediger and his colleagues. In the Roediger paradigm, subjects are
asked to memorize a list of words (e.g., night, dream, bed) related to a target word (e.g., sleep) that was omitted from the presented list. Subjects typically infer the target word and assume that the target word was actually presented. This type of false memory is different from a false memory created to make nonalignable alternatives more alignable. In the latter case, subjects infer a value for a product described on a nonalignable attribute to make two products more comparable, and later assume that this information was actually presented. This type of false memory seems more like an intrusion or a self-generated inference that, once formed, is difficult to discriminate from actually presented information (see Johnson et al. 1993, Psychological Bulletin; see Kardes 1988, Psychology & Marketing, for a source monitoring study conducted using a marketing context). This type of inference is interesting, but different from the type of false memories investigated by Roediger and his colleagues. This type of inference requires schematic knowledge about products, whereas Roediger’s false memories are exhibited by experts and novices. Accountability is also irrelevant for Roediger’s false memories. A schema is a general knowledge structure used for understanding, and it is well known that experts have more well-developed knowledge structures than novices. These knowledge structures or schemas enable experts to form default assumptions about missing information and to infer expectations about what is likely to be true. For examples, experts are likely to infer that a car has four wheels even if this information is omitted from a product description. Because experts have more elaborate knowledge structures, they are also more likely to generate expectations about the likely performance of a product on a missing nonalignable attribute dimension. Schematic knowledge also leads people to distort presented information that is inconsistent with their schemas in a manner that makes this information more schema-consistent.

The coding for the second type of false memory is more problematic. Three different types of recalled information were coded as spontaneous false recalls, even though there are important differences among these three types of recalled information. This category included confusions (confusing information about one brand for another, see Burke and Srull 1988, JCR), new features not presented about either brand, and distortions (misremembered attribute value levels). These types of memories are conceptually distinct and should be treated as conceptually distinct analytically.

In my opinion, the authors overstated the novelty of their contribution by ignoring prior marketing research showing that expertise, involvement, and accountability can increase bias in some specific situations; see


Across the five experiments, false memories had relatively little influence on product evaluations. This diminishes the practical implications of this research, and the reason for these null results is unclear. If false memory incidence rates were extremely low they should not have a strong influence on product evaluations. Using the current coding scheme, some types of false memories seem more relevant to product evaluations than others. Furthermore, a disconnect
between memory and judgment should be expected when judgments are formed in an online manner rather than in a memory-based manner (Hastie and Park 1986, Psychological Review).

Reviewer A argues that Study 1 should be deleted because it contributes little beyond the remaining studies. I agree with this recommendation and assessment. Sobel tests should also be added, and greater justification should be provided for the false memory coding scheme that is used. Reviewer A also noted that memory performance for existing aligned attributes was high regardless of the level of expertise, consistent with prior research on attribute alignability.

Reviewer B also argued that the false memory coding scheme that was used is difficult to justify, and that more than two false memory categories are needed. Reviewer B also noted that the coding scheme used in Study 1 is different from that used in the remaining studies. Reviewer B also asked for a table of the incidence of false memories broken down by the type of false memories and expertise.

Reviewer C offers a more parsimonious explanation for the observed results: experts remember more than novices, experts have more accurate recall than false recall, and accountability depends on the motivation and ability to recall the presented information. Reviewer C also argues that the alignability explanation and the accountability explanation need greater embellishment and justification. It is also difficult to discuss accuracy in Study 4, and it is unclear why memory performance was measured by free recall rather than by cued recall or recognition. Study 5 also needs to be more strongly motivated by a more detailed discussion of process vs. outcome accountability.

Overall, this study appears to examine some interesting issues, but the results are difficult to interpret due to ambiguities present in the current false memory coding scheme. As Reviewer B indicates, the false memory coding scheme that is used drives the results, and the completeness and accuracy of the false memory coding scheme determines what can be learned from this study. In my opinion, more than two categories of false memories are needed, and some of the current subcategories (e.g., confusions) seem less important than others. A more detailed and more carefully justified false memory coding scheme is needed.

**Reviewer A**

**Comments to the Authors:**

This manuscript was a pleasure to read and review. The authors replicate the basic finding that experts are more prone to false recall than novices but significantly expand our understanding of the processes that lead to this effect. Specifically, the paper demonstrates that although differences in schema complexity can explain spontaneous false recall, expertise-based increases in accountability fundamentally drive other forms of false recall (false recall of aligned features in the product comparison tasks at hand). The authors then demonstrate the effects of accountability and these false recall patterns on judgment quality. I was very positively disposed to the presentation of the background literature, the theoretical development, the chosen methodology, and the consistency of the results. I have some suggestions and propose some minor refinements, but my overall evaluation is very positive.
Detailed Comments/Suggestions:

1. I would recommend dropping experiment 1. I find no fault in its construction, but it adds nothing to our understanding that is not already demonstrated in experiment 2. Given that the manuscript currently contains five experiments, I think the removal of experiment 1 will not negatively impact overall contribution and will increase the length to contribution ratio. It will also allow a quicker transition to the testing of hypotheses 3 and 4.

2. I think the authors should explain in more detail their rationale for what constitutes a “spontaneous false recall” and an “attempt-to-align false recall” (page 12). More specifically, the authors argue that a subject who falsely recalls that option A had “No option to play DVD” is demonstrating a spontaneous false recall but a subject who falsely recalls that option A had “No option to play DVD” AND that option B had “Option to play DVD” is demonstrating an “attempt-to-align false recall.” I understand the basic argument here, but this latter attempt to align false recall seems qualitatively different from a false recall in which the subject recalls a feature in option B that aligns with a presented feature of option A. Your results all clearly hinge on this distinction, so I think further explanation/motivation is warranted.

3. Related to point 2 above, I think the authors could use the distinction they are drawing between forms of false recall to indicate whether the false recall results are based in encoding or retrieval. The inference of an aligned feature that was presented for the opposing option could be produced at either encoding or retrieval. However, the “double” false recall effect seems more likely a product of retrieval in which the expert is generating attributes they know products within the category either have or do not have and then falsely remembering which options did or did not possess the attribute. Again, this seems dangerously close to a spontaneous false recall, so a better development of your typology of false recall forms would be useful.

4. Throughout the manuscript, the authors use the Baron and Kenny framework for testing mediation. I have no problem with the Baron and Kenny framework, but would like to see the Sobel Test Statistic reported for each of these mediation tests as well.

5. This is a more minor point, but on page 12, you describe three sources of false recall of aligned features. The first source mentioned situations in which the subject “…recalled the incorrect value of the original option” and the third source mentioned situations in which the subject “recalled a feature value other than that which was originally presented.” These seem identical to me. If they are not identical, please provide more explanation of how they are distinct.

6. Abstract specificity: I would suggest being more direct with your theory and findings here. Rather than saying you “add to this growing research by showing how an additional mechanism contributes…” I would suggest more directly identifying the mechanism and its effects on both false recall and evaluation.
7. In your discussion of the effects of expertise on correct recalls, one curious effect that warrants brief comment is the absence of an effect of expertise on correct recall of existing aligned features. Since the theory suggests that experts are driven to false recall of nonexistent aligned attributes, it seems logical that that bias would also improve rates of correct recall of aligned features. It seems likely that the absence of this effect is due to increased attention to aligned attributes by novices. As a result, the relative increase in correct recall of aligned attributes would not differ for experts and novices. This effect is not central to your theory, but I think a brief analysis of it in your experimental discussion would be helpful.

Well done.

Reviewer B

Comments to the Authors:

The goal of this paper is to determine the circumstances in which experts’ well developed knowledge structures may actually hinder memory and judgment relative to that of novices. The research also sheds light on the underlying mechanism that accounts for these effects. The studies were well designed and executed and the results are interesting and somewhat counter intuitive (this is a good thing). I found the paper to be very well written.

With this said, I have one main concern that makes the interpretation of the results somewhat questionable as well as a few other less major issues.

Main Concern

The results of all five studies hinge on the validity of the coding method that was used to categorize a false memory as being either an instance of “spontaneous false recall” or an instance of an “attempt-to-align false recall.” Since many of the main effects rely on correctly categorizing the false memories, and most of the process evidence relies on dissociations found between these two types of memories, the coding scheme becomes of utmost importance. With this said, there are a number of criteria that were used to categorize false memories as being of one of two types that do not seem to be consistent with the definition of each type.

Spontaneous false recall is defined as a “non-presented associate, p.5” that, due to experts’ advanced knowledge structures, is inadvertently activated, and hence falsely recalled as having been presented. On page 5 an example is given whereby investment-related words were recalled as having been previously presented within a list when in fact they never appeared on the list. However, one of the three codes used to categorize a false memory as being a spontaneous false memory seems to be inconsistent with this definition. The first criteria, recalling the correct value of an aligned feature but for the wrong product option, fails this definition on two fronts. First, it is not a false memory for an item that was never presented, and second, it seems to be much more consistent with source confusion (confusion as to which option had that particular value) as opposed to spontaneous activation as a result of an advanced knowledge structure.
Attempt-to-align false recall is defined as an attempt to align non-aligned features so that “when recalling a feature for one option, experts…falsely recall a comparable counterpart for the other option, p.6.” Hence, appropriately so, a false memory due to an attempt-to-align can only be evident if the feature is correctly recalled for one option AND falsely recalled for the other option. If the feature is not correctly recalled for the option for which it was actually present, but only falsely recalled for the option in which it was not present, then this is simply source confusion, and not an attempt-to-align false memory. Source confusion does not rely on an attempt to align or even an attempt to compare. However, an item was categorized as an instance of an “attempt-to-align false recall” if a non-aligned feature was falsely stated as having been associated with one option when it was in fact associated with the other (see top of page 12). This criteria should have been expended to indicate that the same subject would also have had to have correctly recalled the comparable feature for the alternative option.

It is hard to know how much these potential mis-categorizations impact the results because I have no way of knowing what percentage of recalled items relied on these criteria for categorization purposes. Notably, I do not suggest simply altering the definitions of the two types of false memories simply to fit your coding scheme, as the definitions themselves seem quite reasonable. At a minimum I would like to see the data re-analyzed with those items that relied on the questionable criteria mentioned above being removed from the data set – maybe they should be placed into your third category (those false memories that did not seem to fit into either of these false recall categories).

Less major concerns

I would have liked to have seen the results replicated with another product category but in my mind it is certainly not worth collecting more data to do so.

Is gender confounded with expertise? Intuition would suggest that a much higher percentage of males (versus females) consider themselves experts on video game systems. What happens if you add gender to the analysis?

I would like to see the mean number of false memories for both false recall types across conditions for all studies.

Items were included when analyzing the total number of false recall items in study 1 that were clearly not false memories. Data should be re-analyzed with these items removed (items such as, “I hate video games” and “this is boring” are clearly not false memories).

Reviewer C

Comments to the Authors:

Overview

This paper intends to demonstrate conditions under which experts are more prone to errors of recall. Secondarily it claims that such errors can be exacerbated by experts’ sense of
accountability. Finally, it claims that these recall errors lead to errors in judgmental accuracy. However, as it is currently written and reported it seems possible that a simpler story is available: a) experts remember more than novices; b) because experts have more recall they have more mistaken recall; c) accountability should increase recall but only for those who are involved in the task and capable or retaining memories (i.e. experts).

Big Issues

1) **Do experts just remember more?** It appears so: All five studies report a significant effect of expertise on total recall; experts appear to have better correct recall as well (in Studies 1, 2 and 5; Study 3 not reported; Study 4 marginally significant). I am wondering why we would not expect experts to make a few more mistakes if their overall recall is much higher – if so, this does not seem to be a story about experts being “more likely” to make mistakes. [The gambler that plays the slots more often loses more often too, but this does not mean he is a worse gambler.] Is it possible to analyze the *percent* of recall that is mistaken? Otherwise, the theoretical story presented here is unnecessarily complex.

2) **The alignment story is not as strong as it needs to be.** The distinction between attempt-to-align recall and spontaneous recall errors is potentially interesting but neither the hypothesizing nor the data as it is currently reported bear this out: H1 and H2 suggest precisely the same sort of effects regardless of type of error. Can you build a better story around why expert errors are driven by alignability (rather than by a base rate as suggested in 1) above)? Is there something about comparison processes that particular disfavors experts? What would happen, for example, if participants were asked to recall information about each product independently? Sporadically in the studies there are different-ish results for attempt-to-align and spontaneous error (i.e. Study 2’s mediation results); these are either problems for your theory (if you do not expect different results) or assets (if you do). Please clarify!

3) **The accountability story can also be improved.** The accountability story is tacked onto the end of the recall story almost as an afterthought, but if subsequent analysis really sustained the idea that accountability can undermine recall that would be pretty interesting news. In fact, I think this would make a clearer contribution to the literature than the present positioning: Accountability in general improves attention, involvement, accuracy. Accountability in conjunction with expertise perhaps even more so – a novice bereft of memory structures may not be able to improve memory regardless of accountability [Imagine someone who doesn’t speak Chinese and a native speaker are both set a task of remembering a list of Chinese characters – accountability will not help the nonspeaker much]. So, if you can get consist results showing the inverse interaction (experts get worse with accountability but novices improve), that would be interesting.

4) **Accuracy?** I don’t think we can fairly claim “less accuracy” from Study 4; unless someone actually rates Option A as better than Option B. How can we say that someone who sees Option B as somewhat better than A is incorrect? Without a better accuracy measure the accuracy claims in the abstract and introduction need to be modified or eliminated.

Other Issues
--Why measure expertise subjectively if it is the cognitive structure underlying memory that should drive your effects?

--Why is free recall that best memory measure for your theory?

--Study 5 is not well motivated – no hypothesizing about process v. outcome accountability is done.

--It would really help clarify the paper if you gave some examples of alignability and non-alignability up front. I am left wondering if it is fair to experts to claim an error if they infer “no Bluetooth” for Option A when “Bluetooth” was mentioned for B. True, if no value for Bluetooth was mentioned for A this is an error in the technical sense. But especially given the side-by-side comparison of the alternatives, is it not fair for the respondent to infer “no Bluetooth” if Bluetooth is not mentioned? It seems as though this would be an accurate inference for many ‘real world’ circumstances.

--Results are incompletely reported – I think we need to see all types of recall for all studies. In some places means are not reported in the text, in others the figures are incomplete. Key variables (i.e. Study 4 evaluation results) have neither means reported in the text nor figures, so far as I can see.

**Trainee Reviewer**

**Comments to the Authors:**

This paper proposes that when experts are put in a product comparison context, their higher sense of accountability for their judgments leads to false recall. This is an interesting topic from a theoretical and application perspective.

I share the author’s belief that there should be more examination of false memory in consumer contexts. Additionally, I believe that the authors have found results that advance not only consumer literature but psychological literature. There are ramifications from this research that may be directly usable in contexts other than product comparison. However, I am interested on not only the process of this paper but also WHEN it occurs in the memory process. Does this false memory occur because of accountability at the time of encoding or does it occur because of accountability at the time of retrieval?

The five experiments were very clear and were built directly from the theoretical background. I believe that this is a very strong paper and has interesting ramifications for the product comparison literature and false memory literature.

**Theoretical**

The theoretical perspective comes from several different areas: theories of expertise from consumer behavior and psychology, theories of false recall, and theories of product comparison.
The past research on expert false recall suggests that expert knowledge schemas are so developed that they interfere with information retrieval and lead to false recall. However, the authors identify a key gap in the expert false memory research by identifying unique forms of false expert false memory in a product comparison context. Specifically, authors conclude that experts engage not only in the commonly known spontaneous false recall but also coin a new term for false recall: “attempt to align false recall”. The authors focus on showing expert fallibility in the context of memory based product comparison in terms of recall ability and decision making – well done.

The authors cite past research that when consumers make product comparisons, they compare products either (a) across the same dimensions or (b) compare a feature that is unique to only one of the products. The difference between experts and novices is that novices will tend to engage in easier comparisons (just engage in (a)) and experts will try to make unique features available in only one product align with features in the comparison product. This attempt-to-align process will lead experts to fill in the gaps and lead to false memory. I recommend that the authors try to describe this attempt-to-align process a bit more clearly in the theory development. I was unsure of exactly what you meant or how this could be put into practice until I got to E1 and the DV was explained. Does this process only occur in a product comparison context or could it be extended to other contexts?

The authors theorize that this attempt-to-align false recall stems from the fact that experts have more developed schemas than novices and that experts feel more accountability which will in turn lead to sub-optimal product evaluation. Does the attempt-to-align process occur at the time of encoding or retrieval? Are the expert consumers filling in these gaps when they encode the content or is there something that is happening at the time of retrieval that is causing this process to occur? I think that identifying WHEN the process occurs is key to your argument that feelings of accountability and effortful product comparison are the mechanisms behind attempt-to-align false recall.

Methodological

Overall, the methodological work in this paper was very strong and built directly from the theoretical processes described above.

The first study found support for H1 and H2. Of specific interest was H2 whereby experts showed more attempt-to-align false recall than novices. Looking back to the theory development, I thought that this experiment did a good job of teasing apart the differences between spontaneous false recall and attempt-to-align false recall. Try to communicate these differences earlier in the paper.

The second study found that experts have a higher sense of accountability and this higher sense of accountability causes greater attempt-to-align false recalls. Spontaneous false recalls were not mediated by accountability. This was a strong study and led one step closer to identifying the mechanism behind attempt-to-align false recall in study 3. I really enjoyed study 3 and found it to be the strongest study in the paper. Particularly, I thought that introducing the idea of involvement was very important. Introducing this concept in the theoretical section would have
been particularly strong, especially by showing that there could be competing theories for the proposed process.

The fourth and fifth study examined measurable effects on the product evaluations after false recall. E4 showed that attempt-to-align false recalls lead to suboptimal product evaluations but spontaneous false recalls do not affect product judgment. E5 was key to showing that for experts, it is important to highlight outcome accountability and for novices it is important to highlight process accountability in order to lead to optimal product evaluation.

These five studies were very strong. Study 3 and 5 were particularly strong. I would have enjoyed more theoretical development for both of these studies in order to make for a stronger paper.