Dear Editor,

Thank you very much for giving us the opportunity to submit a revision of Ms BR-Org-22-240 entitled "Introducing a Two Buttons Sets Procedure for Advancing the Study of Retrieval Awareness in Evaluative Conditioning Effects" to BRM. We are grateful for the constructive comments and helpful suggestions provided in the editorial correspondence. As we explain in our response to the reviewers and should be apparent in this revision, we paid careful attention to the various points raised by the reviewers. Of note, we also followed-up upon your and Reviewer 3’s suggestion to collect additional data. The design used for the new study supports our previous conclusions after addressing the concern communicated by Reviewer 3 that our findings may have been artifactually created by the presence of non-paired and valent stimuli in the two previous studies. As an additional asset, the new study allowed for a direct comparison between conclusions derived from our procedure versus the procedure introduced by Waroquier et al. (REF). To foreshadow, the new study replicates our findings but fails to replicates those reported by Waroquier et al. (REF). Although we can only speculate about the reasons for this lack of replication (which we do in the general discussion), the new study firmly supports the main goal of the current endeavor: introducing and validating a new procedure for testing the contribution of various forms of memory to evaluative conditioning effects.

As was the case for the first two studies, the procedure introduced in the new study (i.e., essentially an extension of the TBR procedure to a three-button sets procedure) allows addressing a very large number of questions of theoretical interest for attitude research and research on memory, and accordingly, we have collected additional data to explore these issues. We believe that it would be unreasonable to address all pre-registered theoretical hypothesis as well as all additional exploratory analyses in the context of a single methodological contribution. To repeat, our intention here is to introduce and validate a new procedure. It is not to introduce this new procedure, to validate it, and to use it to address all theoretical questions we were interested in. Doing so would likely result in a manuscript whose size is well over 140 pages (a previous manuscript based on the first two studies already exceeded 100 pages). It would also result in a manuscript that may be considered more a theoretical than a methodological contribution (see also the concern raised by Reviewer 1). We hope that you and the reviewers understand our choice to transparently not follow-up on all pre-registered hypothesis in the context of this single contribution, but to rather follow-up on those hypothesis that were relevant to our current validation endeavor.

Again, we thank you for giving us the opportunity to improve our manuscript. Needless to say, we would be happy to address any remaining point that could help further strengthen our present contribution.

Yours sincerely,

Dear Prof. Stahl:

Manuscript ID BR-Org-22-240 entitled "Introducing a Two Buttons Sets Procedure for Advancing the Study of Retrieval Awareness in Evaluative Conditioning Effects", that you submitted to Behavior Research Methods, has been reviewed.

Thank you for submitting your research to Behavior Research Methods.

I now received three reviews of your submission (see the comments at the bottom of this letter). In addition, I have read the ms myself. Reviewer 1 apologizes for not being able to do an in-depth review as a result of personal circumstances. Still, I think the review is important to be included and I thank all three reviewers for their comments.

As you will see, the reviews are critical, as maybe befits the thorny issue you address. The recommendations are two major revisions and one rejection. Often this would result in a rejection. However, because of the importance of the topic, I am willing to give you a chance to respond to the reviewers' comments. I don't think this can be done properly without including a new study addressing the various points raised.

Please note that my invitation is no guarantee of eventual acceptance. The revisions asked are major and the quality of your revision will need to be assessed again, if possible by the same reviewers. If the quality of the revision is not good enough, the ms is likely to be rejected at that stage. If you find the required revisions too onerous, you are, of course, free to submit the ms to another journal.

To revise your manuscript, log into <https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fmc.manuscriptcentral.com%2Fbrmic&amp;data=05%7C01%7Colivier.corneille%40uclouvain.be%7C86857bd8bd924ccde60d08da6000899e%7C7ab090d4fa2e4ecfbc7c4127b4d582ec%7C0%7C0%7C637927852742136212%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&amp;sdata=MBzEvfvG6WVCtJsp4QqPUjjY4e0sVz72VgOqgq2UNgs%3D&amp;reserved=0> and enter your Author Center. You will find your manuscript title listed under "Manuscripts with Decisions."  Under "Actions" click on "Create a Revision." The manuscript number will be appended to denote a revision.

You will not be able to make your revisions on the documents previously submitted.  Instead, revise your manuscript using a word processing program and upload it again through your Author Center. Highlight the changes you made in the manuscript by using bold or colored text, preferably in blue rather than red. Do not use track changes or colored background that make your ms difficult to read.

Also include a response to the editor and the reviewers. Do this in the author's response field, not as a separate cover letter or attached file. In this way, your response is available in the reviewer center making it easy for reviewers to find it. Detail the changes you made or why you object to the proposed modifications.

Delete any redundant files (such as the earlier cover letter, manuscript, or figures no longer the same) before completing the submission of your revision.

IMPORTANT: Please note that if you choose to re-submit your article this must be done within 8 months of the date of this e:mail. Submission after this date will be considered to be a new submission, unless you have a good reason why you cannot meet the deadline. Then contact me before the deadline and I will try to give you an extension.

## Reply to Editor

Reviewer(s)' Comments to Author:

## Review 1

Reviewer: 1

Comments to the Author

In general, I found this paper would be of interest to many attitudes and EC researchers like myself.

We appreciate the reviewer‘s assessment of the broad relevance of our research.

However, I found three limitations:

1) conceptually, retrieval awareness (a continuous, real-time processing) and retrieval outcome (a binary, end product) are distinctive processes. The task here captures retrieval outcome, but not retrieval awareness as claimed in the paper. To measure retrieval awareness or real-time retrieval, the authors need to consider EEG or MEG these real-time neural activity and look for memory retrieval-related neural activity.

We are thankful for this comment, which helps us clarify our claims. We agree that our task taps the outcome, not the process dynamics of memory retrieval. In a simplified model, the memory retrieval process can be either successful or unsuccessful; in the first case, the outcome of memory retrieval is US valence, which is reported using the Memory button set; in the second case, nothing can be retrieved from memory, and participants report their attitude using the Attitude button set. Before this background, we use „retrieval awareness“ to refer to the subjective success (or lack thereof) of memory retrieval (i.e., participants‘ assessment of whether they retrieved something from memory or not) – it serves to inform the theorerical issue of whether participants need to consciously retrieve US valence in order to show an EC effect. The term was coined with the intention to contrast this issue from that of encoding awareness (i.e., do participants need awareness of the critical information at encoding to show an EC effect?).

2) The manuscript is mainly about deliberate evaluations using direct measures. The results and conclusions may not readily generalize to spontaneous evaluations captured by indirect measures. Given the theoretical and practical significance of spontaneous evaluations, I would find the present manuscript not sufficiently impactful to the field.

We agree that the paper focuses on direct evaluative measures, and readily concede that indirect methods may yield other results. However, we maintain that direct measures are an important part of the picture, especially so when subjective first-person assessments of awareness are in the focus (as in the present case). In addition, from an applied perspective, direct subjective assessments are typically the only means available to people when they rely on their attitudes to make decisions in everyday situations.

3) I also found the paper is less a method paper. To me, this paper is more a project addressing a theoretical question, with a new task. Thus, it is not clear to me whether the paper fits the journal’s aim.

While we agree that our paper also addresses a theoretical question with a new task, we maintain that the paper focuses on the method more than on the theoretical issue, and that the goal of this paper is to establish the new task’s validity. To us, it seemed impossible to discuss the new task independent of its theoretical background. We hope, therefore, that this does not preclude publication in BRM.

## Review 2

Reviewer: 2

Comments to the Author

The paper proposes a new task (TBS) aimed at studying whether EC may occur without consciously recollecting US valence (being aware of remembering or being aware of knowing) and the influence of attitudes on valence identification (the role of affect-as-information).

The procedure aims to better respect the immediacy and information criteria (e.g., Shanks & St. John, 1994) than the previous approaches (PDP, item level analyses).

Participants were exposed to an EC procedure in which each CS was paired with several US of either positive or negative valence. During the subsequent TBS task, CS as well as positive and negative foils were presented. Participants had four response options on a single screen. If they remembered that the stimulus had been paired with a US they indicated US valence (positive vs. negative). If they don’t, they reported their attitude toward the stimuli (positive vs. negative).

The authors found that, when participants reported remembering US valence they correctly identified US valence in 70% to 80% of the trials. When they reported not remembering US valence, there was no EC effect (i.e., pairings had no significant effect on binary CS evaluations). They also found that initial attitudes toward foils or CS (in study 2) influence binary CS evaluations and Memory-set responses.

I found the research question addressed by the paper important. Most of the previous literature has focused on the role of awareness during encoding. The role of awareness during the evaluation stage also seems important to me. It is linked to the question of EC controllability as ‘retrieval awareness’ may be viewed as a precondition to control one’s evaluative responses. I found the new method interesting. In particular, it satisfies the immediacy criteria better than previous trial-by-trial approaches. I describe below the different points that prevent me from recommending accepting the paper in its current form.

1) Being able to use information intentionally (i.e., strategic control) is one of the properties of consciousness captured by the PDP (e.g., intentional vs. unintentional). Awareness of knowing is another feature associated with consciousness that is captured by subjective measures of awareness (e.g., Dienes & Seth, 2010; Timmermans & Cleeremans, 2015). My understanding is that the TBS capture the second aspect. According to the authors, the TBS aims at making the best of ‘by item’ and PDP approaches (see p7) at the methodological level. I suggest discussing more explicitly which aspect(s) of consciousness (i.e., theoretical level) is or may be assessed by the TBS.

We thank the reviewer for pointing out this distinction, which we discuss in the revised manuscript (GD, section Limitations, first subsection, pasted below).

In our understanding, the previous by-item measures of valence memory required participant's judgment knowledge (i.e., whether a CS had been paired with positive or negative USs), as well as their awareness of knowing (i.e., it asks participants to respond 'dont know' if you don't have knowledge). The TBS similarly taps these two aspects of awareness, in a slightly different way (i.e., report your attitude if you don't have knowledge).

In the absence of awareness of knowing, it might be possible to control the use of unconscious information strategically - but this, in our view, does not characterize the PD task as it is applied to EC by Hütter et al. We would argue that the PD approach proposed by Hütter et al. builds on the same awareness of knowledge as the TBS: To be able to apply the relevant part of the instruction, participants had to decide whether they have knowledge or not . This is also in line with Waroquier et al who have characterized the PD task as a subjective measure.

In our understanding, having awareness of information implies the ability to use that information intentionally (if not necessarily on all measures, if those measures have additional requirements). In that sense, both the TBS and PD tasks tap awareness of knowing.

For your convenience, the section is pasted below:

"Intentional control versus awareness of knowing

When focusing on the consciousness feature of (un)intentional control, it can be debated whether the TBS task is indeed an improvement upon the PD approach. Tapping the ability to intentionally control information is critical in many PD applications, which typically aim at dissociating two distinct contributions to behavior, only one of which is under intentional control. The inclusion-exclusion instruction manipulates how the controllable process is mapped onto the set of responses, and it is assumed that the uncontrollable process affects responses in the same way in both instruction conditions. In other words, participants in the PD task can only reverse the response assignment in the exclusion condition for intentionally controllable information; the response pattern remains unaffected by the instruction manipulation insofar as there are also uncontrollable contributions.

On the other hand, by-item measures of valence memory required participant's judgment knowledge (i.e., whether a CS had been paired with positive or negative USs), as well as their awareness of knowing (i.e., they ask participants to respond 'dont know' if they are unaware of any knowledge). The TBS similarly taps these two aspects of awareness, albeit in a slightly different way (i.e., it asks participants to report their attitude if they are not aware of having any knowledge). For both the by-item and TBS tasks, we assume that the subjective knowledge state (i.e., being aware of either having knowledge or of not having knowledge) can be used intentionally (i.e., participants intentionally select whether to report their memory or their attitude).

Importantly, controlling the use of a piece of information or knowledge is not the same as being aware of knowing it (i.e., control may be possible for unconscious knowledge, that is, for which awareness of knowing is lacking; it is also conceivable, at least, that awareness of knowledge may be present in some situations while that knowledge uncontrollably affects a specific task). Hence, if the TBS and the PD approach (as applied to assessing EC without memory by Hütter et al.) were to capture distinct features of consciousness, this might explain different patterns of results (and would prohibit presenting the TBS task as an improvement of the PD approach).

While it may generally be possible to control the use of unconscious information strategically (i.e., in the absence of awareness of knowing), we would argue that this does not characterize the PD task as it is applied to EC by Hütter et al. In our view, participants in the PD approach also requires awareness of knowing as well as judgment knowledge in order to follow instructions: To be able to select the relevant part of the instruction (i.e., report memories or attitudes), participants have to decide whether, subjectively, they have knowledge or not (i.e., they need awareness of their knowledge); to be able to report their attitude (or memory), they require judgment knowledge (i.e., whether a CS had been paired with positive or negative USs). Importantly, this knowledge must be under their intentional control, so that they can follow the respective inclusion or exclusion instruction (i.e., report veridical versus reversed attitudes or memories). In this sense, the PD approach to assessing EC without memory is not a traditional PD paradigm that capitalizes on (un)controllability of (in this case) attitudinal information --- to the contrary, it assumes and requires attitudinal information to be controllable. Instead, it hinges on the subjective assessment of having (versus not having) conscious knowledge about CS-US pairings; it is this subjective awareness of (not) knowing that determines whether a response is reversed or not.

^[This interpretation is in line with Waroquier et al. () who have also characterized the PD task as a subjective measure.]"

2) Related to the previous point, I suggest including a discussion of other studies, that used a trial-by-trial measures, which are relevant to the ‘retrieval awareness’ question. For example, Jurchiș et al. (2020) conditioned positively strings of a first artificial grammar and negatively strings of a second grammar. During a test phase they assessed participants liking of new strings from both grammars and found an EC effect. The EC effect was found in trials in which participants reported having no idea what makes the string likeable (i.e., no awareness of knowing). Another study that used a trial-by-trial approach found that both intentional and unintended use of memory contributes to EC (Halbeisen et al., 2014).

Thank you for pointing out these studies. We agree that a more in-depth treatment of the substantive matter is desirable but had refrained from doing so in the initial submission because of the journal's focus on research methods (rather than the substantive issue of retrieval awareness). We have extended our review of the relevant literature in the revision---although only slightly so, because we feel that a fair and thorough discussion of the cited works and their relation to the present study is beyond the scope of the article.

The revision now refers to the Jurchis et al. finding in the GD section, as an example for research in related fields: "Finally, recent findings in the literature on artifical grammar learning (AGL) have found evidence for EC in the absence of awareness of knowing (Jurchiș et al., 2020). In that study, awareness of knowing and EC were assessed using distinct measures; hence, the result is potentially susceptible to a lack of immediacy. It would be interesting to test whether the finding can be replicated in a variant of the TBS task applied to the AGL paradigm."

3) In the current paradigm, each CS is paired with several US. It doesn't seem obvious to me what is retrieved when a participant indicates ‘positive’ or ‘negative’ in a Memory-set response (the same question can be raised for several previous studies). They could recollect a case where a specific CS was associated with a specific US (conscious recollection). They could remember that they experienced a positive/negative feeling during one of the CS-US presentations. They might notice that there are regularities in the different pairs they are exposed to and remember this observation later. What is retrieved can influence participants’ metacognitive states (e.g., remember, know, guess…) and their responses to the TBS task.

Clearly, different scenarios are possible for retrieval of knowledge about the US's valence - but note that this is also the case when each CS is paired with only one specific US. We have collected additional memory measures to explore the specifics of the retrieved information as well as the metacognitive states, TODO and report some of our findings in the revised manuscript. Retrieval of a specific CS-US pairing instance would render a successful identification of the specific US in an "US identity memory" task more likely; and it would be likely associated with "remember" reports; retrieval of a positive or negative gist memory may be associated with lower accuracy of US identification and perhaps more "know" reports. Future research should use more appropriate methods to separate cases of detailed retrieval from those involving gist retrieval, and investigate associated EC effects.

In the revision, Experiment 3 used a single-US pairing, yielding the same findings as in the previous two studies. While there may be differences in retrieval experiences, the result pattern across studies is identical, suggesting that the type of pairing does not affect the validity of the TBS.

4) The TBS provides memory data when participants report they remember the pairings and binary evaluations when they report they don’t. So we don’t have the same information in the two cases. Participants also evaluated CS on scales. Is there a significant EC effect on evaluative ratings when there is ‘retrieval awareness’ and no effect when there is no retrieval awareness? The authors also took some memory measures for all CS. Is memory performance above chance when there is and when there is no ‘retrieval awareness’? This is relevant from a theoretical perspective, as in other incidental learning paradigms, several experiments suggest that participants acquire knowledge about the material even if they are not aware of acquiring it (e.g., Dienes & Scott, 2005; Paciorek & Williams, 2015). As there are two steps (learning the associations and ratings) in EC, it is possible that participants remember the pairings and there is no EC effect or there is no effect because they did not learn anything.

Indeed, the Ec effects on the rating scales are in line with the findings frmo the TBS. There is EC only for Memory cases, not for Attitude cases.

TODO: We would be happy to report some of the additional data to address memory accuracy for M vs. A - although please note that our studies only allow for tentative explorative first impressions, rather than firm conclusions. To avoid an overly long article, we collected these additional results in an online appendix.

Please note that, with respect to the theoretical assumptions about the learning processes, we are not aware of models assuming two learning steps, and hence perhaps cannot address this comment in the article. In our reading of the literature, there are S-S and S-R models, both assuming the creation of a single association (between either CS and US, or between CS and UR).

5) The authors did not find any significant EC effect in the absence of ‘retrieval awareness’. On its own, this result does not provide evidence that EC do not occur in the absence of ‘retrieval awareness’. Interpreting <i>p</i>-values does not indicate whether the results are not significant because of a lack of power or because there is no effect in the population (e.g., Dienes, 2014). Previous, research found small EC effects in the absence of ‘retrieval awareness’ (e.g., Mierop, Hütter, Stahl, & Corneille, 2018; Waroquier et al., 2020).  To assess the extent to which the data are more likely under H<sub>0</sub> than under the hypothesis of a small effect, a Bayesian analyses would be needed. Interpreting the CI around effect size (reported in the paper) could also be an option (Cumming, 2014). i.e., a narrow CI around zero indicates that the population effect is close to 0 with a good confidence.

Thank you for pointing this out. Of course, as the present paper focuses on presenting the TBS method, the issue of whether the effect is zero is not central. However, we believe that (in addition to addressing some methodoogical concerns), the results from Experiment 3 provide strong evidence for the absence of EC-without-memory. We hope that these additional data address your concern.

6) Information may be available but not recollected. So there are two interpretations if EC cannot be found in the absence of ‘retrieval awareness’ in the current paradigm. 1) EC requires that people consciously recollect the pairings or US valence when making their evaluations. E.g., An advertisement is only influential if people think about it when they shop. 2) In the TBS paradigm, participants are explicitly asked if they remember US valence which encourages them to try to remember the pairings.  If no explicit instruction is given, the pairings might not be recollected (thought it could be recollected) but still influence evaluations. In a related paradigm, Hütter & Sweldens (2018), study 4 found that the estimation of controllable processes, measured via a PDP-type approach, was zero in an incidental learning condition (covert instruction) but much larger when participants were instructed to apply US valence (inclusion) or the opposite valence (exclusion). Moreover, EC effects have proven to be quite robust even after a delay (e.g., Hütter et al, 2012). In this case, do people still recollect the pairings when making their evaluation?  In the current version the authors only consider interpretation 1. The different possibilities could be discussed.

Thank you for pointing out this subtle yet important distinction of two cases of retrieval awareness. The second possibility is now discussed in the Limitations subsection on "Deliberate memory retrieval and evaluations" (see below); we explicitly leave open the possibility that an EC effect may obtain without memory retrieval. This will require challenging yet very interesting future research.

The relevant paragraph reads: "Relatedly, the TBS explicitly asks participants whether they have memory for the CS-US pairings; this instruction should induce them to try to retrieve their memory for the pairings. What we found is that EC is only observed with successful retrieval. However, it is possible that participants would show EC effects without retrieving CS-US pairing memory, if they were not explicitly asked to do so. In other words, whereas EC may require pairing memory, it may not require retrieval of those memories at the time of CS evaluations. This theoretical possibility could be tested, for instance, with the help of indirect measures of memory retrieval in future studies."

7) The paper shows that pre-conditioning attitudes have an impact on memory responses when participants report remembering US valence (effect of CS valence on memory set responses). Does attitudes acquired through affect misattribution (i.e., participants mistakenly attribute their affective response to the CS instead of attributing it to the US during encoding, Jones, et al., 2009) could lead participants to mistakenly believe that they remember US valence and to indicate the correct valence in the memory response set in the absence of memory? Is it possible to control /to evaluate this potential bias with the new procedure?  This could perhaps be done by taking into account US identity memory.

Todo: discuss/explore. One way to investigate this is to compare the effect of CS valence on memory-set responses across different levels of US valence: If misattribution is the underlying process, it should be CS valence, not US valence, that determines responses. Additionally taking into account the accuracy of US identity memory may help: Misattribution would predict inaccurate US identifications when CS and US valence differ. Note that in our data, this could not be detected because we presented participants only with USs of the correct valence as response options on the US identity memory task. To investigate this issue, additional data would therefore be needed.

*… does R2 interpret our findings appropriately? Does the misattr. account make sense?*

8) I missed an interpretation of the effect of CS pre-attitude on attitude change.

Todo: discuss.

## Review 3

Reviewer: 3

Comments to the Author

I read this paper with great interest. The authors pursue an applaudable goal, namely separating the cognitive foundations of evaluative conditioning effects, one based on explicit memory for the valence of the unconditioned stimuli one based on an implicit form of memory that leaves people with a vague positive or negative feeling towards the conditioned stimuli in the absence of explicit memory for the source of this feeling. This question is relevant to the debate on awareness that has been going on for decades, but has still not resulted in strong, conclusive evidence.

The authors present a combination of previous methodological approaches and two experiments aimed at validating their approach that they term the “two-button-sets” (TBS) procedure. The main advantage of this approach is that it does not require an exclusion condition as the original PD procedure introduced by Hütter and colleagues (2012). This can be considered a simplification.

However, this simplification comes at costs, which I summarize in two conceptual concerns. I also have a set of methodological concerns resulting from the design of the experiments.

We are thankful that the reviewer deems the removal of the exclusion condition as a welcome simplification. The TBS does not require an exclusion condition, which is a procedural simplification, but it is also a simplification of the underlying data-analytic model and its core assumptions. In particular, the model no longer needs to assume that processes are identical across conditions (i.e., the invariance assumption) data are independent across participants; both assumptions have often been found to be violated in similar MPT models of related domains. Hence, if the reviewer’s conceptual concerns were to invalidate other advantages oft he TBS, and this were the only benefit of the TBS, we would argue that it would still be a worthwhile improvement over previous methods.

Conceptual concerns

(1) Do the two response sets indeed constitute a single measurement?

The authors claim that their procedure is a significant advantage over the previous per-item approach, because it allows assessing memory and evaluation in a single measure. However, given that the measure involves two questions and two sets of response options, I am not convinced participants approach the tasks as a single measure. Do these two sets of questions constitute a single measure, just because they are presented on a single page? I am not sure. The two response sets can be rephrased as two sets of “pleasant” – “unpleasant” – “skip” responses, whereby the use of the skip option depends on the suitability of the other question. That is, there is indeed a strong dependency between the two measures. Is this the reason why the authors claim this to be a single measure?

 „(1) Do the two response sets indeed constitute a single measurement?“ Perhaps not perfectly, but they approximate it better than separate measurements administered blockwise (or even on two separate screens). Even if there are two memory retrievals, the TBS minimizes measruement error by holding the temporal context constant. The dependency issue raised by the reviewer is another point: If construed as two separate questions, they must be answered together (so even if two retrievals, then close in time). To contrast the scenario sketched by R3, we can present a theoretically way in which global-matching memory models would be able to address this is in a single retrieval (cuing memory with the CS retrieves an echo strength – that informs the decision whether or not I „remember“ US valence; and an echo content – that informs whether the context associated with that CS was more likely positive or negative; i.e., as we modeled EC in the reversed-extinction paper).

(2) Are the measures of memory and attitude really process-pure?

The authors seem to assume that the inclusion of a second, evaluative response set prevents participants from using their evaluation of the CSs as a (valid) cue towards the paired US valence (i.e., the first response set). The authors own data contradict this claim: Affect-as-information is demonstrated to influence memory responses in both experiments. I was confused as to how this finding constitutes a validation of the TBS. Isn’t a more appropriate interpretation that affect-as-information cannot be prevented by the TBS? A similar concern arises for the inclusion of the ”I don’t remember” option. The inclusion of this response option may not prevent affect-as-information as suggested by the current data as well as by the reversed EC effects obtained by Stahl and colleagues (2009). Given that participants know that their responses can be evaluated for accuracy, they might still be tempted to guess based on cues they deem informative.

 „(2) Are the measures of memory and attitude really process-pure?“ No, and we never claim it (whereas the PDP does). We can explain more explicitly that the TBS does not assume purity, which is another advantage over the PD. In addition, our data of course show that purity is violated (which has never been tested for the PDP; but given the similarity of tasks, it is likely also violated there). The TBS cannot prevent the \*use\* of an affect-as-info heuristic, true. But it avoids a \*biasing effect\* of the use of this heuristic on conclusions about EC-without-memory (as the PDP aims to do; but its assumptions do not consider affect-as-info for „memory“ responses). And the TBS can \*detect\* the use of this heuristic (by using nonpaired valent stimuli), which is another benefit over previous tasks which do not use those stimuli.

Methodological concerns

(3) Does the experimental design warrant strong conclusions?

A major methodological concern relates to the experimental design of the studies presented to validate the TBS. Specifically, the authors include stimuli that were never paired. In Experiment 1, participants’ task is thus not only more complex than the task in the original PDP presented by Hütter and colleagues (2012), but this complexity is also not matched by the response options given to participants. Specifically, the procedure introduces another level of meta-memory, namely knowing whether or not a stimulus was shown during the conditioning phase. Participants have no chance to indicate that they know that a stimulus was not shown, and they receive no information on how to deal with stimuli for which they know that they were not paired. I can only speculate about the effects that such a misleading task set may have. Possibilities are that participants feel inclined to use the memory response set, because they are supposed to use the attitude one only if they do not remember the status of the conditioned stimulus, or that they choose a random response on either of the two response sets as none of these apply to their memory state.

We agree that, compared to some previous studies, the composition of the stimulus list in our studies was more complex. Yet, we do not think that this enhanced complexity necessarily extended to participants‘ task, which we believe is certainly less complex than the exclusion task required by the PD approach.

In the TBS, for a given CS, participants were asked to remember and report the associated US’s valence, or (if they don’t remember) report their attitude towards it.

For non-paired stimuli, participants cannot retrieve a memory about US valence. Hence, in this case, they use the attitude response set to report their attitude towards the nonpaired stimulus.

In the manuscript it is stated that participant’s understanding of the instructions was probed, but the results are not reported (or I did not find them). Did participants indeed have a clear understanding of the instructions, or did they have difficulties mapping their mnemonic states onto the response sets?

TODO: We dealt with this issue in the same way as Hütter et al. Have addressed it: Participants were tested on their understanding of instructions, and if they failed, they were shown the instructions again. Participants failing after the 3rd time were excluded from analyses… ??? The results showed that a substantial proportion of participants had difficulty understanding the instructions. This is, however, also true for other tasks (e.g., the RKG task). Most importantly, our empirical tests of instruction understanding ensured that those participants who passed the test were well able to perform the task as instructed.

Another potential problem stems from the combination of neutral and extremely valent stimuli within participants. This approach will likely affect the interpretation and use of “pleasant” and “unpleasant” response options. Presumably, the difference between extremely positive and extremely negative conditioned stimuli is more extreme than the difference between positively and negatively conditioned stimuli. In the present context, the latter will appear to provide a much less clear signal than the former. This concern also affects the assessment of the EC-without-memory-effect. This analysis probes for an influence of US valence over and above an influence of (the pre-existing) CS valence. However, the relatively weaker signal resulting from the unconditioned stimuli may not become apparent in this particular setting.

This concern also applies to previous studies using the PD approach, as well as the studies by Waroquier et al. (), who have also shown extremely valent stimuli during the pre-conditioning ratings. In these studies, evidence for „unconscious“ EC has been found despite the presence of extremely valent stimuli, suggesting that their presence is not problematic.

Summary

I agree with the authors that the demonstrable contribution of EC effects in the absence of explicit memory for the pairings in the original PD procedure is very small and that the significance of this finding (especially in applied settings) can be debated. I also agree that the previous methods, the PD procedure and the per-item approach have limitations. Although I also agree that the authors’ approach addresses some of the limitations of earlier approaches, the authors did not convince me that the current research presents an improvement over the per-item method or the PD procedure or convincing evidence against the role of EC effects in the absence of memory. The authors might be onto something useful, but the set-up of the task and the validation approach require further development.

We are grateful for the challenge of further validating the TBS method. We conducted another study that addresses the reviewer’s methodological concerns regarding the inclusion of items that were never paired. In this study, only paired items were included in the TBS task. The results were identical: There was again no EC in the absence of memory.

Minor issues

- On p. 18 the authors state that US identity memory was at chance, but those data were not presented in the paper.

Thank you; this is corrected in the revision.

- It is suboptimal to refer to non-paired stimuli as conditioned stimuli.

Thank you; this is corrected in the revision.

- I was confused as to why the PD approach does not meet the information criterion. The authors seem to suggest that this is due to the combination of two conditions, the inclusion and exclusion condition, but I could not derive information as to why this should affect the information criterion in the opinion of the authors.

Thank you for pointing this out; it is a side-effect of minimizing our discussion of the PD approach in the present paper because it would distract from its main goal. The information criterion highlights the need for tapping exactly those pieces of information that could be used to consciously produce an EC effect. The EC effect is about a specific CS-US pair presented to a specific person; so the information needs to be collected on an items-within-persons level. This is not the case in the PD task, which aggregates data across not only items and persons, but also across the inclusion and exclusion conditions.

- The preregistration documents on OSF differ quite a bit from the present paper, both in terms of the narrative and the set of analyses. The authors also included measures of confidence, memory for US valence, and memory for US identity intended for correlational analyses with responses obtained with the TBS, which are not reported in the manuscript. I do not want to say that all analyses must be conducted and reported as planned (I’m actually not sure how informative these correlational analyses would be). However, I think it is necessary to explain in more detail why the report in the manuscript differs from the initial goals and interpretations. The current manuscript reads as if the studies were analyzed and reported exactly as preregistered with the one exception that the exclusion condition was omitted.

Thank you for pointing this out. The preregistration reflects our plan to do two things at once: validate the TBS, and probe the validity of Hütter er al.‘s PD approach. A previous version of the manuscript, which presented the entire set of results as planned, was deemed overly long and complex by reviewers (it can be accessed as a previous version of this manuscript on PsyArxiv). We therefore split the results and, in the present paper, focus on validating the TBS (before turning to the second goal in another paper). The narrative and set of analyses needed to be adapted to this new way of presenting the results.

We regret if our current way of presenting the studies comes across as misleading in any way. We certainly did not aim at misrepresenting our study or hiding any results (as can be verified, e.g., by the fact that the previous version of the manuscript is available on PsyArxiv). We simply aimed for the most straightforward way to describe the differences between the planned studies and the reported data. If there is a better ways to more transparently describe the situatioon, we are happy to revise our manuscript accordingly.