**Reviewer #1:** Summary: In this study, the authors compared the influence of predictable vs. random task sequences on the size of global switch costs (measures as the performance difference between pure blocks and non-switch trials in switching blocks) and local switch costs (measured as performance difference between non-switch trials and switch trials in switching blocks). The results showed such an influence as global switch costs were larger with predictable than with random task sequences, while local costs were larger with random than with predictable task sequences. The authors attributed these differences to different working memory demands as predictable task switching requires participants to engage in keeping track of the sequence, which is not necessary for a random task sequence.  
  
Evaluation: I agree with the authors that there is relatively little research directly comparing the size of switch costs in random (i.e., cued) vs. predictable (i.e., alternating runs) switching. Thus, such a systematic approach as presented in the present study is certainly helpful to better understand the cognitive control mechanisms that are at place in these different task-switching paradigms. However, in order to make an outstanding theoretical and empirical contribution to the existing literature, the authors of this study still need to provide more clarity regarding the theoretical motivation of their study and especially the expectations they formulate (please see my major comments 5 - 7).  
  
**Major comments:**

1) Terminology/Task: The authors refer to the task they use in the present experiment as the Consonant-Vowel/Odd-Even task (CVOE) task. More specifically, in their description the CVOE task is not only specific for a certain type of task combination (i.e., the combination of a consonant-vowel and an odd-even task) but also for a specific combination of pure and mixed blocks (see page 5, line 4: "this task presents participants with pure and switch blocks, the CVOE task allows for computation of local and global switch costs."). I have to admit that it did not know the CVOE task as a specific paradigm to measure both local and global switch costs before. The combination of a consonant-vowel and an odd-even task was already used in relatively early papers on task switching (cf. Rogers & Monsell, 1995) so this is not a novel task combination. Also, the combination of block types varied between the present study and the study by Huff et al. (2015) as only an alternating runs-switching block was used in the later study. Finally, other tasks (e.g. a combination of magnitude and parity tasks) could be used to measure local and global switch costs within one experiment - which would then not correspond to an CVOE task. Thus, taken together, I am not sure whether using the term CVOE task to express both a specific task combination and a specific block type combination is the best way to go. However, please not that this is mainly a matter of terminology. Both the task combination and the paradigm of the study are perfectly fine to address the research question at hand.  
  
2) Introduction: In the introduction, I was wondering why the authors first focussed so much on the Stroop task as this was not part of the current study. This is also connected to the fact that the authors distinguish between task-relevant and "highly salient but task-unrelated" information at the beginning of the introduction. Such salient distractors certainly occur in case of the Stroop task. However, in task switching, and especially in the current CVOE task, I would argue that one stimulus (letter vs. number) is not more salient than the other. This is especially true as the authors later on argue that both tasks in this CVOE setting are comparable (see page 5, line 2f: "An advantage of the CVOE task is that it assesses task-switching performance using a balanced design in which both tasks are equivalent in difficulty"). My suggestion here would be to limit the introduction somewhat more specifically to the present study.  
  
3) Introduction, Page 2: The authors argue that the Stroop task is specifically used to explore task-set inhibition and that this is a difference to task-switching paradigms that rather relate to attentional control and working memory. This distinction is most justified if only task-set reconfiguration is considered as an explanation for local switch costs (which is indeed the case in the introduction of the present study, see page 4, first paragraph). However, there are other explanations for the emergence of local switch costs in which task-set inhibition plays a major role. In task-set inertia accounts (e.g., Allport et al., 2014), one main assumption is that currently irrelevant tasks become inhibited and a part of switch costs is due to overcoming this inhibition when a task is needed again, thus, when one switches back to a recently inhibited task. In my opinion, it would be important to also consider task-set inertia accounts in order to provide a more complete picture (as the authors already do in the discussion on page 21).  
[Reference: Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), Conscious and nonconscious information processing: Attention and performance XV (pp. 421- 452). MIT Press.]  
  
4) Page 4, line 7 from below: The authors discuss the different impact between univalent and bivalent stimuli and describe the difference as "participants must keep both task-sets active in working memory and, prior to responding, must quickly consider which response corresponds to the correct task-set on each trial". In my opinion, keeping two task sets active in working memory is mainly what distinguishes pure blocks from switching blocks (i.e., even when switching between two tasks with univalent stimuli, one has to keep two task-sets active in working memory). Bivalent stimuli, however, may additionally activate the non-relevant task set on a bottom-up manner as the currently irrelevant stimulus dimension is presented nevertheless. Thus, the activation of the competing task sets when using bivalent rather than univalent stimuli may be based more on such a bottom-up mechanisms than on the requirement of keeping two tasks active in working memory.  
  
5) If I read the Procedure correctly, a task-cue was presented in all switching blocks (i.e., in both random and in predictable, alternating-runs blocks). Additionally, participants were not informed about the alternating sequence of tasks in the respective block (page 13, last line "they received no prior instructions regarding the specific sequence for each switch block"). This means that the difference between the random and the predictable blocks were minimized as participants did not know about the sequence and were not required to keep it in memory (as the cue was always presented). Maybe this was done to make both types of switching blocks as comparable as possible. Yet, this also raises two questions. First, is there any evidence that participants indeed monitored the sequence in alternating-runs blocks? If they did not (as it was neither required nor instructed), how can the difference between block types then be attributed to differences in the working memory demand? And second, both switching blocks now differ from pure blocks, in which no cue was presented. Thus, the difference between pure blocks and non-switch trials could also be (partially) attributed to cue encoding processes.  
  
6) The aim of the authors was to compare random and predictable switching. To do so both local switch costs and global switch costs were used as empirical measures. When using switch costs as empirical measure, it is always important to keep in mind that they are a relative measure as we do not have an a priori baseline. For example, local switch costs can increase due to faster responses in non-switch trials OR due to slower responses in the switch trials (or both). Similarly, we can express switch costs as costs (due to task-set reconfiguration processes, for example) or we could rather focus on task-repetition benefits (due to persisting activation of the relevant task). The authors do acknowledge this problem at many places in their manuscript. Still, I had sometimes had difficulties to fully understand their theoretical explanations about when to expect smaller or larger switch costs. For example, on page 5, line 5 from below the authors argue that task-set reconfiguration should be more problematic in random switching "as the unpredictable nature of random sequencing should be particularly taxing for working memory processes relative to predictive alternating runs". At this point I would suggest to more explicitly explain, why this should be the case (is it due to the ability to prepare for a task switch in predictable switching that is missing in random switching?) and which specific effect or process should be more difficult in random than in predictable switching. In general, I believe that it would be helpful to more explicitly state whether an assumed process affects non-switch and/or switch trials in a specific type of switching.

**In my opinion, there are three arguments that play a critical role here:**

**First**: Task predictability may lead to a better performance in switch trials as participants can expect a task-switch and potentially, they also can start task-set reconfiguration already before the next stimulus is provided so that especially switch trials in a predictable switching block should benefit. Is this what the authors mean be "unpredictable switch trials are especially taxing when participants must reconfigure task sets" (page 20). In general, the authors could add some ideas and literature on task-set preparation here to make this point more explicit.

**Second**: Previous studies demonstrated that non-switch RTs get even faster in longer runs of task repetitions ◊ this should affect mainly the RTs in non-switch trials in random switching as only here longer runs of task repetitions can be expected. Additionally, one might also argue that after a longer run of task repetitions, it becomes harder to switch the task (which could be argued based on task-set inertia assuming an increased task activation for the currently relevant but an increased inhibition for the currently irrelevant task). In this way also switch trials in random switching might be affected.

**Third**: In predictable task switching, participants have to keep the sequence in memory and does have to perform an additional task which might affect (i.e., reduce) the performance in both non-switch and switch trials in predictable switching.  
  
7) Minar and Shah (2008) observed higher local switch costs for predictable vs. random switching and higher global costs for random vs. predictable switching. Thus, the data pattern is completely the opposite to that of the present study. However, this discrepancy is only very briefly mentioned at a very late point the general discussion. Further, it is mainly attributed to differences in the methodology (including a difference in the number of participants). As there is very little research that examines the size of switch costs in predictable and random switching (as the authors themselves emphasis), I would expect that such a discrepancy in findings is discusses to a larger extend and on a more theoretical level.  
  
Minor comments:  
  
- Page 3, line 4: "using a rule on one subset of stimuli but switching to a different rule when cued" - this formulation could be misinterpreted as having a constant stimulus to task mapping so that one specific subset of stimuli (e.g., letter-digit pairs containing vowels) would only be used in the consonant-vowel task. There are studies that compare such constant vs. varied stimulus-task mappings but this is not the case in the present study. Thus, I suggest to use a different wording to avoid confusion.

- Stimulus material: Is there a specific reason to use two-digit numbers? For an odd-even decision, only the number at the unit position is relevant, the decade is irrelevant for the odd-even decision and rather may leadto an effect of decade-unit parity (in)congruity (for a review see Nuerk et al., 2011). More specifically, a parity decision might also take place for the decade so that 48 is congruent (both decade and unit signal even) while 47 is incongruent (the unit indicates odd whereas the decade is even). One reason might be that two-digit numbers allow for a larger set of stimuli. Yet, this is probably even more problematic as then the numbers than letters are used. This could in the extreme case lead to differential learning effects and working memory demands (remembering a stimulus-response mappings might differ between numbers and letters). [Reference: Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2015). Extending the mental number line. Zeitschrift für Psychologie, 219(1), 3-22. <https://doi.org/10.1027/2151-2604/a000041>]

- Page 15, exclusion of trials: For RT analyses only correct trials were included. I was wondering whether trials following an error were included in RT and error analyses as these might be difficult to be categorized as switch or non-switch trials (i.e., when participants committed an error because they performed the wrong task on the previous trial, the following trial might be a non-switch rather than a switch trial or vice versa).

- At the beginning of the General Discussion (and also at other places in the manuscript as in the beginning of the conclusion), the authors argue that they explored "the effects of predictive and random sequenced task-switching on working memory and attentional control". I was wondering whether the present study can indeed tell us something about how the type of sequencing impacts working memory. The study shows influences on the size of local and global switch costs, which then could be explained by working memory processes. However, there is no direct measure (or dependent variable) that directly taps into working memory. Thus, I believe such a statement is a bit misleading.

- Page 20, line 10f: "Random switching led to greater local switch costs, while predictive switching led to greater global switch costs." - For this sentence I would suggest to provide also the information what the respective comparison is. More specifically, when reading this sentence "random switching led to greater local switch costs" this could be interpreted as "greater than predictive switching" or "greater than global switch costs". It is relatively clear from the context what the comparison is. Still, a clarification could be helpful.

- Table 1: I suggest to reverse the order of non-switch and switch trials as non-switch trials are compared to pure blocks on the one hand and switch trials on the other hand. So, at least for me, non-switch trials are "in the middle".  
  
  
Andrea M. Philipp  
(signed)  
  
  
  
Reviewer #2: The present paper investigates differences in task-switch costs between two task-switching paradigm: a random switching condition (in which tasks occur in a random sequence) and an alternate runs condition (in which tasks vary according to AABBAABB). The results indicate larger local switch costs (switch vs. non-switch trials) for the random switching condition but larger global switch costs for the alternate runs condition (non-switch trials vs. trials from a single-task condition). These effects are interpreted in terms of working memory and attentional control demand.  
  
The question addressed in this paper is reasonable and interesting for the community. However, I have major concerns regarding the theoretical framing and the interpretation of results, as explained in more detail below.  
  
1.  
The introduction is overly long (11 pages) but still fails to theoretically motivate the study and derive reasonable hypotheses:

- The authors provide a lengthy description of their own previous studies on aging and Alzheimer dementia, but it is unclear how this is related to the current study, in which only healthy young participants are included. It appears that they aim to argue for a relationship between global switch costs and working memory. However, showing that participants with impaired working memory have impaired global switch costs does not necessarily imply that global switch costs reflect working memory load. For instance, Mayr (2001) showed global switch costs while controlling for working memory demands, and argued that the costs reflect control processes rather than working memory load. The authors cite this paper but largely neglect it in their reasoning. Actually, the authors ignore most relevant literature that has not been published by their own group.

- The main question of the study is whether global and local switch costs differ between an alternating runs condition and a random task switching condition. However, despite the lengthy introduction, this question is not really motivated by theoretical ideas. On pp. 6-8, the authors only describe which studies have been done and which comparisons/conditions have not been implemented yet, and this alone appears to motivate the study. Later on p.10/11, the empirical predictions come out of the blue and are either unspecific (increased local switch costs for random switching because "the unpredictable nature of random sequencing should be particularly taxing for working memory processes" on p.10) or are unrelated to any of the previously described issues ("participants would be particularly impacted whenever switching occurred at non-predictive intervals, as the lack of discernable pattern would prevent expectancies of upcoming trials" or "alternating-runs sequence also requires that participants attend to the position of each trial within the sequence").  
- The distributional analyses are also not really motivated by theoretical considerations. On p. 9, the authors state "when task-switching, tau would be expected to increase whenever switching places additional strain on attentional control systems. Thus, tau would be expected to show an increase for random rather than predictive switching." Again, this prediction is too unspecific as it is unclear why more attentional control demand is imposed on the random switching condition.  
In sum, the paper lacks a clear theoretical framing and a sufficient justification of the hypotheses and predictions. I recommend substantial rewriting of the introduction. Hypotheses and predictions must be derived from an adequate discussion of the state of the literature. Studies that are not directly related to the main question should be omitted (or only briefly mentioned). The introduction should be shortened drastically (by about 50%).  
  
2.  
The methods section is unstructured and unclear in many ways. For instance:  
- The power analysis is unclear. Why is power reported for main effects given that ANOVAs with interaction effects were used?

- p. 12: The description of stimuli is unclear. For instance, "the list of randomly generated consonants was split in half, such that half were paired with odd numbers, while the remaining half were paired with even numbers." What means randomly generated consonants? Does that mean that some consonants were only combined with odd numbers and others were only combined with even numbers? (That wouldn't make sense at all). Were different stimuli generated for each participant? It's also unclear how many stimuli were generated in this way.

- Information on instructions, timing and trial numbers are distributed across the whole section. This arbitrary switching between levels (trial, block, instruction, etc.) makes it very difficult to follow the description.  
  
3.  
As mentioned on p. 20, the different patterns of local and global switch costs in the two conditions can mainly be attributed to a difference in the RT of non-switch trials. Because global switch costs are calculated as [non-switch minus pure] and local switch costs are calculated as [switch minus non-switch], the increased RT on non-switch trials for the alternate runs condition leads to both, reduced local switch costs but increased global switch costs in this condition. The observation that no difference between conditions is obtained for switch trials provides further support for the idea that the effects in switch costs are due to this shift in non-switch trials (and speaks against the authors' interpretation that the effect on local switch costs is due to a higher demand on switch trials in the random switching condition). This has obviously been mentioned by a reviewer on a previous submission but the response of the authors on p. 20 is not clear to me.

The RT effect on non-switch trials can be perfectly explained by a confound: The alternate runs condition includes only one non-switch (or repetition) trial in a row whereas the random switching condition involved multiple consecutive repetitions of the task (which makes it more similar to the pure block trials). It could be these additional repetitions that produce the faster non-switch trials in the random switching condition, and thus, the effects in local and global switch costs. The authors appear to agree with this, as they mention this interpretation on p. 21. However, they fail to draw the correct conclusion: If the results are due to this confound, one cannot unequivocally infer that the effects on local and global switch costs reflect differences in working memory or attentional control. In other words, the authors' central conclusions are not justified.

One way to address this issue would be to analyze only those trials from the random switching condition that are non-switch trials following a switch trial or switch trials following a non-switch trials (i.e., trials with the same preceding trial category as in the alternate runs condition). However, this still leaves open whether the obtained results are due to differences in the n-2 trial (which could be due to backward inhibition effects, see Mayr & Keele, 2000, JEPG).