

Distorted estimates of implicit and explicit learning in applications of the
process-dissociation procedure to the SRT task

Christoph Stahl, Marius Barth, & Hilder Haider

University of Cologne

Author note

This work was funded by Deutsche Forschungsgemeinschaft grants STA-1269/1-1 and HA-5447/8-1. Correspondence should be addressed to Christoph Stahl, Herbert-Lewin-Str. 2, D-50931 Köln, Germany, Email: christoph.stahl@uni-koeln.de.

Abstract

We investigated potential biases affecting the validity of the process-dissociation (PD) procedure when applied to sequence learning. Participants were or were not exposed to a serial reaction time task (SRTT) with two types of pseudo-random materials. Afterwards, participants worked on a free or cued generation task under inclusion and exclusion instructions. Results showed that pre-experimental response tendencies, non-associative learning of location frequencies, and the usage of cue locations introduced bias to PD estimates. These biases may lead to erroneous conclusions regarding the presence of implicit and explicit knowledge. Potential remedies for these problems are discussed.

Keywords: implicit learning, serial reaction time task, process-dissociation procedure, response bias

Distorted estimates of implicit and explicit learning in applications of the
process-dissociation procedure to the SRT task

Implicit learning refers to the ability to adapt to regularities inherent in the environment in the absence of conscious awareness about the ongoing learning process itself or about the outcome of what is learned. This ability is fundamental for human beings as it allows us to act optimally in stable environments with relatively little effort.

One of the most frequently utilized paradigms in the field of implicit learning is the serial reaction time task (SRTT) originating from Nissen and Bullemer (1987). In this standard SRTT, participants respond to locations on the screen which are mapped to spatially corresponding keys. Participants are instructed to press the appropriate response key whenever an asterisk occurs at a certain screen location. Unbeknownst to the participants, the locations of the asterisk follow a regular sequence. After several blocks of practice, the sequence is replaced by either a new but also regular sequence, or by a random sequence. In this transfer block, performance shows a decrement that disappears almost immediately when the original regularity is reintroduced, reflecting learning of the regularity. Importantly, participants are not able to explicate their acquired knowledge when asked to do so. Even with more sensitive tests including the recently introduced wagering task (Dienes & Seth, 2010; Haider, Eichler, & Lange, 2011; Persaud, McLeod, & Cowey, 2007) or the process-dissociation procedure (Destrebecqz & Cleeremans, 2001; Haider et al., 2011; Jacoby, 1991), explicit knowledge of the sequence is rare. This dissociation between performance and expressible knowledge is generally assumed to indicate implicit learning.

A central defining feature of explicit knowledge is that it is controllable, whereas implicit knowledge is thought not to be under conscious control. Destrebecqz and Cleeremans (2001) utilized this distinction and applied the process-dissociation (PD) procedure to measure sequence learning. In process dissociation, performance in two conditions of the same task is contrasted: An inclusion condition, in which explicit and implicit knowledge both produce the same response, and an exclusion condition in which

explicit and implicit knowledge produce opposing responses (Buchner, Steffens, & Rothkegel, 1998; for applications of the PD to sequence learning with the recognition task see Buchner, Steffens, Rothkegel, & Erdfelder, 1997).

In their application to sequence learning, Destrebecqz and Cleeremans (2001) used a generation task: After SRTT training, participants are asked to generate a sequence of responses that is either as similar as possible to the learned sequence (in the inclusion condition), or a sequence as dissimilar as possible (exclusion condition). To the degree that explicit knowledge is available, the proportion of generated responses that match the learned sequence should differ between inclusion and exclusion. To the degree that implicit knowledge is available, the proportion of matching responses in the exclusion condition should be greater than a chance baseline or control condition. In one group (i.e., RSI = 0 ms), participants were better than chance in their ability to reproduce the regularity in the sequence, even under exclusion instructions (i.e., performance under exclusion condition, E , was above a chance baseline B , $E > B$), a finding that was interpreted as reflecting sequence knowledge. However, performance under inclusion (I) and exclusion instructions was identical (i.e., $I = E$), a finding that is interpreted as indicating the absence of explicit knowledge and instead suggests that the sequence knowledge was fully implicit.¹

The PD procedure is a simple and elegant way to disentangle controllable and uncontrollable processes, which has been widely used across a wide range of research questions (Yonelinas & Jacoby, 2012) and has the potential to address many open questions in the domain of implicit learning and memory. However, some authors have raised concerns suggesting that the assumptions underlying the PD may sometimes turn out to be overly simplified, which, in turn, would threaten the validity of PD results (e.g., Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Curran & Hintzman, 1995; Hirshman, 1998; Rouder, Lu, Morey, Sun, & Speckman, 2008).

For instance, it has been argued that the automatic process may be confounded with extra-experimental influences (Buchner et al., 1995; Rouder et al., 2008) or response

tendencies (Stahl & Degner, 2007). In sequence learning, for example, participants may bring pre-experimental knowledge to the lab that interacts with task properties, or they may attempt to strategically generate non-regular or random sequences especially under exclusion conditions (e.g., Boyer, Destrebecqz, & Cleeremans, 2005). These participants may be influenced by their subjective theories about randomness, which may thereby affect generation performance, and perhaps distort PD estimates of implicit and explicit knowledge.

The type of task may furthermore affect the validity of PD estimates. In applications to the SRTT, discrepancies have been observed between free and cued versions of the generation task (Destrebecqz & Cleeremans, 2001; Wilkinson & Shanks, 2004): Whereas Destrebecqz & Cleeremans (2001), using free generation, have obtained evidence for implicit knowledge (i.e., $E > B$), Wilkinson & Shanks (2004), using a cued generation task, report the absence of implicit knowledge, with exclusion performance not distinguishable from baseline. However, as argued by Q. Fu, Dienes, & Fu (2010), the failure to replicate the $E > B$ pattern could be due to a lower sensitivity of the cued generation task.

Furthermore, different types of random control conditions have been used that differ with regard to simple frequency information or other sequence-unrelated properties (Reed & Johnson, 1994; Stadler, 1992). For instance, participants who are trained on randomly selected permutations of a fixed-length sequence might learn that the entire set of response positions is used up before any position is repeated (negative recency), whereas participants confronted with fully random material during learning might learn that response positions are independent (Boyer et al., 2005). Although this type of knowledge is unspecific and sequence-unrelated, it may nevertheless affect performance in the generation task and distort conclusion about sequence knowledge if not properly controlled for.

Goal of the present study

To gauge these threats to the validity of the PD in sequence learning, and to examine the conditions under which PD can help to understand the processes underlying sequence

learning, we investigated the possibility of differential effects of material (permuted vs. random), task format (free vs. cued generation), and response tendencies on generation performance under inclusion and exclusion conditions.

Simple frequency information and sequence-unrelated properties.

Generation performance may be contaminated by knowledge unrelated to the sequence such as the frequency of repetitions and reversals (Reed & Johnson, 1994). Repetitions and reversals are often not included into the regular sequence because they are considered especially salient and prone to lead to explicit knowledge (Stadler, 1992). If participants (explicitly or implicitly) pick up this absence of reversals in the learning phase, they may use this knowledge in the generation task, for instance, to generate fewer reversals across both conditions, or even to generate fewer reversals in the inclusion than the exclusion task. In other words, if the sequence does not contain reversals, then a generated sequence that reflects this property of few reversals is scored as above-chance performance. The finding that generation performance is above baseline will suggest the presence of implicit sequence knowledge. If the strength of this effect furthermore differs between the inclusion and exclusion instructions, it may artificially produce an $I > E$ finding and lead to erroneous conclusions suggesting the presence of explicit sequence knowledge.

Another type of sequence-unrelated information that may affect generation performance is the frequency of response locations: When some responses are more frequent than others in the learning phase – for instance, in mixed first-/second-order sequences – participants may pick up this information (explicitly or implicitly) and use it in the generation task (Reed & Johnson, 1994). More importantly, if the PD instruction – inclusion or exclusion – can affect the expression of this knowledge, generation performance may be differentially affected, results may be distorted or artifactual, and substantive conclusions might be erroneous.

The present study investigates the degree to which learning of such sequence-unspecific properties (reversals, zero-order frequencies) may distort estimates of explicit and implicit

second-order sequence knowledge.

Free and cued generation tasks. In applications of the PD to sequence learning, two variants of the generation task have been used: free generation (Destrebecqz & Cleeremans, 2001) and cued generation (Wilkinson & Shanks, 2004). In the free generation task, participants are asked to generate a longer stretch of responses without interruption (e.g., 96 trials in Destrebecqz & Cleeremans, 2001). In cued generation, on the other hand, on each trial, a small sequence of stimuli and responses is given as a cue by the experimenter, after which the participant is asked to generate the response that would occur next in the sequence. The results obtained with both tasks have been found to diverge: Using free generation task, Destrebecqz and Cleeremans (2001) reported evidence for implicit knowledge (i.e., exclusion performance was above baseline). In contrast, Wilkinson and Shanks (2004) could not replicate this $E > B$ finding using a cued generation task. The type of generation task may explain this discrepancy if cued generation artificially lowers (or free generation artificially boosts) exclusion performance (but see Q. Fu, Fu, & Dienes, 2008, for a reward-based explanation). Alternatively, the failure to find $E > B$ in cued generation may be taken as evidence for the suboptimal sensitivity of the cued generation task. The present study compares free and cued generation tasks and their ability to detect learning of sequence-unspecific properties.

Response tendencies and subjective intuitions about randomness. In a control condition of an SRTT study, sequence information is typically not available (e.g., Haider et al., 2011). Participants are nevertheless asked to generate transitions reflecting some ‘regular’ sequence under inclusion conditions, and to avoid generating such a ‘regular’ sequence under exclusion conditions. In this situation, subjective notions of regularity and randomness are likely used to generate what participants regard as more regular sequences under inclusion conditions and less regular (or more random) sequences under exclusion conditions. If these subjective notions deviate systematically from the researcher’s notion of randomness that is used to determine the chance baseline, they may bias the pattern of

results and distort substantive conclusions. The present research addresses the question whether response tendencies may bias generation performance, whether response tendencies are acquired during the learning phase or reflect pre-experimental biases, and whether such a bias differentially affects inclusion versus exclusion conditions.

Interactions between these factors. The factors discussed above may occur in combination, with the effect of creating potentially more complex distorting influences on PD estimates. For instance, on top of subjective notions of randomness, sequence-unspecific properties of the training materials may be used to inform participants' response tendencies and perhaps lead to systematic differences between inclusion and exclusion. In addition, response tendencies informed by frequency information may interact with variants of the generation task. For instance, if participants have a tendency to avoid generating recent response locations (Boyer et al., 2005), the simple frequency information about reversals or high- versus low-frequency response location may differentially affect performance under different generation task variants (free or cued). The present study explores such potential interactive effects of simple frequency information, task format, and response tendencies.

The current study

In the present study, which is part of a project aiming at evaluating the validity of the PD procedure in sequence learning, we were interested in the ability of the PD procedure to signal the absence of (both implicit and explicit) sequence knowledge when such knowledge is in fact absent. In addition, we wanted to identify appropriate control conditions to serve as a baseline with which to compare experimental conditions.

We aimed at exploring effects of response tendencies, simple frequency information and unspecific properties of the material, and task properties as well as their interactions. Focus of the present study is their potential of distorting PD estimates of implicit and explicit learning. The present study realized three different 'control' conditions without any sequence information:

- a training phase with randomly drawn permutations of a second-order 8-item sequence,
- a training phase with randomly drawn response locations (from a uniform distribution),
- a no-learning condition in which participants merely familiarized themselves with the task.

Orthogonally, we implemented the two different versions of the generation task (free vs. cued).

We investigated whether inclusion and exclusion performance differed under the three control conditions and in the free versus cued task variants. If the PD model yields valid measures of implicit and explicit sequence knowledge, generation performance should be at chance level in all conditions. If, however, simple frequency properties of learning materials also affects generation performance, we would expect this to be reflected in the permuted condition when compared to the random condition. To the degree that response tendencies lead to deviations from chance level, this should be evident from the no-learning condition but reflected in all three conditions.

Method

Design

The study realized a 3 (*material*: permuted vs. random vs. no-learning) \times 2 (*generation task*: free vs. cued generation) \times 2 (*PD instruction*: inclusion vs. exclusion) \times 2 (*block order*: inclusion first vs. exclusion first) design with repeated measures on the instruction factor.

Participants

190 participants (143 women, with a mean age of 25 years, range 18-59 years) completed the study (data from 8 participants could not be used due to a programming error, one participant failed to follow instructions during SRTT). Most were undergraduates

from University of Cologne. Participants received either course credit or 7 Euro for their participation and were randomly assigned to experimental conditions.

Materials

We used two different types of pseudo-random material:

- A *permuted* sequence was randomly generated for each participant anew by drawing with replacement from the set of all possible permutations of a second-order 8-item sequence. For a given participant, each sequence was construed by randomly selecting two of the six locations to occur twice in the sequence. Per block, 18 permutations were drawn from the set.
- A *random* sequence was randomly generated for each participant anew by drawing with replacement from a uniform distribution of six response locations.

In a third *no-learning* condition, participants performed 20 responses drawn randomly in order to familiarize themselves with the task. They were instructed, prior to the generation task, to imagine they had just worked on a learning phase and to generate the sequence they may have encountered there.

In all conditions, the sequence adhered to the following (additional) restrictions: (1) there were no direct repetitions of response locations, and (2) there were no response location reversals (i.e., A-B-A). As a consequence of the random generation process, frequencies of response locations, first-order transitions, and second-order transitions varied across participants. To determine correct responses in the generation task, we computed an individual criterion for each participant based on their individual transition frequencies.

Procedure

The experiment consisted of three consecutive parts: First, participants worked on a SRTT (the *training phase*), followed by a *generation task* and a postexperimental interview.

In the learning phase, participants in the permuted and random conditions performed a SRTT consisting of 6 blocks with 144 trials each (total of 864 responses). Participants in the no-learning condition performed only 20 random trials to familiarize themselves with the SRTT. SRTT and generation task were run on 17" CRT monitors with a screen resolution of 1024px \times 768px. The viewing distance was approximately 60cm. A horizontal sequence of six white squares (56px \times 56px) was presented on a grey screen. The distance between squares was 112px. Each screen location corresponded to a key on a QWERTZ keyboard (from left to right Y, X, C, B, N, M). Participants had to respond whenever a square's color changed from white to red by pressing the corresponding key. They were instructed to place the left ring-, middle- and index fingers on the keys Y, X and C. The right index-, middle- and ring fingers were to be placed on keys B, N and M. There was no time limit for responses in the learning phase (nor in the generation phase). A warning beep indicated an incorrect response. The response-stimulus interval (RSI) was 250 ms.

Following the SRTT phase, participants were told that stimulus locations during the SRTT followed some underlying sequential structure (participants who were not exposed to the SRTT phase were asked to imagine that they had experienced an SRTT in which locations followed some underlying sequential structure). The generation instructions were presented next, with order of inclusion vs. exclusion task counterbalanced across participants. Under inclusion (exclusion) instructions, participants were told to generate a sequence that is as similar (dissimilar) as possible to the sequence from the learning phase. For both instructions, participants were instructed to follow their intuition if they had no explicit knowledge about the underlying sequence. Direct repetitions were explicitly discouraged and were followed by a warning beep.

In the *free* generation task, after an initial sequence of three cue locations, participants freely generated 120 consecutive response locations. Participants were instructed that three squares would appear to which they had to respond; subsequently, question marks appeared at all locations and participants' key presses were reflected by the corresponding square's

color changing to black. In the *cued* generation task, in each of 120 trials, 3 to 5 stimulus locations (taken from learning materials) were presented as cues, and participants had to respond with the corresponding key, in order to activate any sequence knowledge, after which the next response location had to be generated by the participant. Participants were instructed that a few squares would first appear to which they had to respond; subsequently, the question marks appeared and participants were asked to freely choose the trial's final response location.

Upon completing the computerized task, participants were asked to complete a debriefing questionnaire containing the following items: 'Did you notice anything special during the task? Please note everything that comes to mind.', 'One of the tasks mentioned a sequence in which the squares lit up during the first part of the study. In one of the experimental conditions, the squares did indeed follow a specific sequence. Do you think you were in this condition or not?', 'How confident are you (in %)?', 'Can you describe the sequence in detail?'. Subsequently, participants were asked to indicate, for each of the six response keys, the next key in the sequence on a printed keyboard layout. Finally, participants were thanked and debriefed.

Results

For all analyses, a significance criterion of $\alpha = .05$ was used. If sphericity was violated in repeated-measures ANOVAs, Greenhouse-Geiser-corrected degrees of freedom and p values were used. The raw data (as well as additional supplemental materials) are available from <https://github.com/methexp/pdl1>.²

SRTT Training phase

The mean RTs obtained over training blocks are shown in the left panel of Figure 1, separately for permuted and random material (in this and all following training-phase analyses, we excluded RTs of the first trial of each block and of trials that resulted in an error). Analysing RTs using a 2 (*material*: permuted vs. random) \times 6 (*block number*)

ANOVA revealed a significant main effect of *material*, $F(1, 117) = 5.62$, $p = .019$, $\eta_G^2 = .042$, with slower responses for random material than for permuted material, and a significant main effect of *block number*, $F(2.93, 342.59) = 13.69$, $p < .001$, $\eta_G^2 = .011$, reflecting practice effects. The factors *material* and *block number* did not interact, $F(2.93, 342.59) = 2.02$, $p = .112$, $\eta_G^2 = .002$, indicating that any knowledge was acquired already during the first block.

For the group with permuted material, some locations were presented twice within each pseudo-sequence. The right panel of Figure 1 shows the mean RTs for these two types of stimuli. A 2 (*frequency*: high vs. low) \times 6 (*block number*) ANOVA revealed a main effect of *frequency*, $F(1, 55) = 19.31$, $p < .001$, $\eta_G^2 = .030$; high-frequency responses were faster than low-frequency responses. The main effect of *block number* was also significant, $F(3.72, 204.69) = 6.13$, $p < .001$, $\eta_G^2 = .014$, but not the interaction, $F(3.87, 212.88) = 1.80$, $p = .132$, $\eta_G^2 = .001$. These findings suggest that participants in the *permuted* condition acquired some form of simple response-location frequency knowledge that benefitted their SRTT performance, and especially so for the more frequent responses.

Generation task

In the free generation condition, after an initial sequence of three cue trials, participants freely generated 120 consecutive response locations. In the cued generation condition, in each of 120 trials, three to five stimulus locations (taken from learning materials) were presented as cues and participants had to respond with the corresponding key. For each of the 120 trials, a response *triplet* consisted of the previous two locations as well as the location of the current response (in the cued generation condition, the response triplet consisted of the last two cue locations and the current response location; in the free generation condition, the response triplet consisted of the previous two response locations as well as the current response location; for the first two trials of each generation block, the locations of the corresponding cue trials were used). We calculated the proportion of triplets that were consistent with training sequences (see Appendix A).³ Figure 2 depicts the pattern

of correctly generated triplets as a function of generation task, material, and PD instruction.

Correctly generated second order transitions. The proportions of correctly generated triplets were analysed using a 3 (*material*: permuted vs. random vs. no-learning) \times 2 (*generation task*: free vs. cued) \times 2 (*PD instruction*: inclusion vs. exclusion) \times 2 (*order*: inclusion first vs. exclusion first) ANOVA. It revealed a main effect of *PD instruction*, $F(1, 169) = 4.75$, $p = .031$, $\eta_G^2 = .010$, more correct triplets were generated during inclusion than during exclusion blocks. This $I > E$ pattern suggests the presence of explicit knowledge, despite the absence of sequence information in the training material.

The ANOVA also revealed an interaction of *material* \times *generation task*, $F(2, 169) = 5.88$, $p = .003$, $\eta_G^2 = .042$. Analysing only free generation using a 3 (*material*: permuted vs. random vs. no-learning) \times 2 (*PD instruction*: inclusion vs. exclusion) \times 2 (*order*: inclusion first vs. exclusion first) ANOVA revealed a main effect of *material*, $F(2, 83) = 6.72$, $p = .002$, $\eta_G^2 = .076$ (and again the main effect of *PD instruction*, $F(1, 83) = 4.13$, $p = .045$, $\eta_G^2 = .024$). Tukey's HSDs revealed a significant difference between *permuted* and *random*, $p = .001$, the difference between *permuted* and *no-learning* group trended to be significant, $p = .083$, *random* and *no-learning* groups did not differ from each other, $p = .257$. Analysing only cued generation revealed no effects, suggesting that the free generation task was more sensitive in picking up learning effects on generation performance.

To summarize, there was a main effect of PD instruction as well as an effect of material in the free-generation data; no effects were obtained in the cued-generation data. In the process-dissociation logic, the effect of PD instruction suggests the presence of explicit knowledge. In addition, the effect of material suggests the presence of implicit knowledge in the permuted condition. Next, we will apply two different PD analysis strategies to these data to investigate more formally whether, given the assumptions underlying the PD approach can be upheld, the pattern is indeed indicative of the presence of explicit and implicit knowledge.

Ordinal PD. We first applied the ordinal PD approach (Hirshman, 2004). It proposes a model-free analysis strategy of PD data by deriving four critical data patterns of

inclusion and exclusion performance changes across two conditions that imply differences in automatic and controlled processes between the conditions. For instance, it can be concluded that both the automatic and the controlled processes are stronger in one condition if inclusion performance is greater in that condition but exclusion performance is comparable in both conditions. Also, it can be concluded that the automatic process is greater in one condition if both inclusion and exclusion performance are greater in that condition.

We screened the free-generation data for these data patterns by separately comparing inclusion and exclusion performance across conditions of the material factor. Inclusion performance was affected by material, $F(2, 83) = 5.42, p = .006, \eta_G^2 = .116$: it was greater in the *permuted* than in both the *no-learning* group, $F(1, 55) = 6.70, p = .012, \eta_G^2 = .109$, and the *random* group, $F(1, 53) = 11.65, p = .001, \eta_G^2 = .180$; in the latter two, performance was comparable, $F(1, 58) = 0.08, p = .777, \eta_G^2 = .001$. Exclusion performance was also affected by material, $F(2, 83) = 2.93, p = .059, \eta_G^2 = .066$; specifically, it was greater in the *permuted* than the *random* condition, $F(1, 53) = 5.26, p = .026, \eta_G^2 = .090$, but did not differ between the other two conditions (*no-learning* vs. *permuted*: $F(1, 55) = 0.24, p = .628, \eta_G^2 = .004$; *no-learning* vs. *random*: $F(1, 58) = 3.53, p = .065, \eta_G^2 = .057$).

Comparing the *permuted* and *no-learning* conditions, performance was greater in the *permuted* group under inclusion instructions but was identical in both groups under exclusion instructions. This data pattern implies increased levels of both the automatic and the controlled process in the *permuted* as compared to the *no-learning* condition.

Comparing the *permuted* and *random* conditions, performance was greater in the *permuted* group under both inclusion and exclusion instructions. This data pattern implies an increase in the automatic process in the *permuted* as compared to the *random* condition but no effect on controlled process.

These results suggest that participants acquire some form of implicit knowledge from *permuted* material that they can use to produce above-chance levels of correct triplets in the free generation task. They also suggest that the *permuted* and *random* materials allow

participants to acquire some form of explicit knowledge which they can use to perform better under inclusion than under exclusion conditions.

PD equations. The same interpretation was suggested when the PD equations were used to obtain quantitative estimates of explicit and implicit knowledge. Based on the opposition logic, correct performance under inclusion instructions can arise due to controlled processes (C) or – should these fail – due to automatic processes (A), $p(\text{correct}|\text{inclusion}) = C + (1 - C) * A$. Under exclusion instructions, the controlled process leads to an incorrect response, and correct performance is based only on automatic processes in the absence of controlled processing, $p(\text{correct}|\text{exclusion}) = (1 - C) * A$. Automatic and controlled parameters could differ across experimental conditions, resulting in a total of 12 parameters (i.e., separate sets of A and C for two generation task variants, multiplied by three levels of the material factor). Parameters were estimated using the HMMTree software and the MPTinR package (Singmann & Kellen, 2013; Stahl & Klauer, 2007). Parameter estimates are given in Appendix B.

Hypotheses are tested by imposing restrictions on parameters and evaluating whether these restrictions significantly harm the model’s goodness of fit, in which case the associated hypothesis cannot be upheld. ⁴

To investigate the evidence for any form of learning, we analysed whether there was evidence for explicit knowledge by testing whether controlled parameters could be restricted to zero. This was not possible without substantially harming goodness of fit, $\Delta G^2_{(df=6)} = 28.19, p < .001$. This is consistent with the above findings of an effect of PD instruction, as well as the conclusions drawn from the ordinal PD approach about the presence of explicit knowledge. As above, this effect was strongest in the free generation condition $\Delta G^2_{(df=3)} = 23.55, p < .001$; it was clearly present in both the permuted ($\Delta G^2_{(df=1)} = 13.95, p < .001$) and the random material ($\Delta G^2_{(df=1)} = 9.60, p = .001$). In the cued-generation condition, there was a much weaker effect which was only marginally significant, $\Delta G^2_{(df=3)} = 4.64, p = .074$, and restricted to the permuted material

($\Delta G^2_{(df=1)} = 2.03, p = .077$) but absent for the random material ($\Delta G^2_{(df=1)} = 0.00, p > .999$).

Next we analysed whether there was evidence for implicit knowledge, first by testing whether a restriction of A parameters to an a-priori chance level of .2 can be maintained. This was clearly not the case, $\Delta G^2_{(df=6)} = 172.68, p < .001$. Even if only the no-learning condition was considered, the A parameters were greater than the a-priori chance level, $\Delta G^2_{(df=2)} = 74.26, p < .001$, suggesting that participants had acquired substantial implicit knowledge. The A parameters could also not be set equal across materials, $\Delta G^2_{(df=4)} = 53.14, p < .001$, confirming the above findings that the amount of acquired implicit knowledge differed across conditions. This was the case in the free-generation condition, $\Delta G^2_{(df=2)} = 35.92, p < .001$, as well as in the cued-generation condition, $\Delta G^2_{(df=2)} = 17.22, p < .001$. However, the patterns differed across conditions: In the free-generation task, implicit knowledge estimates were ordered *permuted* > *no-learning* > *random*; in the cued-generation task, the order was *no-learning* = *random* > *permuted*.

Taken together, when the PD equations were used to obtain estimates of explicit and implicit knowledge from the present data, results suggest that participants acquired both implicit and explicit knowledge in the control conditions. Whereas explicit knowledge was detected only in the free-generation task, implicit knowledge was detected in both tasks (but patterns differed between tasks).

Interim summary

In three different control conditions, we computed the proportion of generated responses that matched the learning materials to test for any effect of implicit or explicit knowledge acquired from the learning phase. We obtained converging evidence from three different approaches: (1) In ANOVAs, the proportions of correctly generated responses differed as a function of material (this effect was restricted to the free generation task), as well as of PD instruction. (2) The ordinal PD approach, when applied to the free generation data, yielded evidence for greater implicit knowledge in the permuted than in the random

and no-learning groups. It also yielded evidence for explicit knowledge in the permuted condition (and, by implication, in the random condition). (3) The PD equations yielded estimates of the controlled process, reflecting explicit knowledge, that were significantly different from zero for the free-generation data. They also yielded estimates of the automatic process, reflecting implicit knowledge, that were above chance levels for 4 out of 6 experimental conditions.

Taken together, despite the fact that the material in the learning phase did not contain any sequence information, the PD approach using generation tasks yielded ‘evidence’ for both implicit and explicit knowledge. In the following sections, we will try to account for these findings in terms of sequence-unrelated frequency properties, response tendencies, and cueing artifacts.

Sequence-unrelated frequency properties

First, we analysed the proportion of reversals as well as the proportion of high- versus low-frequency locations (as manipulated in the permuted condition) that participants generated in both versions of the task.

Reversals. Figure 3 shows participants’ proportions of reversals (e.g., 1-3-1) generated during inclusion and exclusion (where triplets that contained a repetition were excluded from analyses). By chance, a reversal would be generated in 1 out of 5 cases.

We conducted a 3 (*material*: permuted, random, no-learning) \times 2 (*generation task*: free vs. cued) \times 2 (*PD instruction*: inclusion vs. exclusion) \times 2 (*order*: inclusion first vs. exclusion first) ANOVA of the proportion of reversals. The ANOVA revealed a main effect of *PD instruction*, $F(1, 169) = 22.14$, $p < .001$, $\eta_G^2 = .057$: More reversals were generated under exclusion than under inclusion instruction.

The ANOVA also revealed a main effect of *material* $F(2, 169) = 3.75$, $p = .025$, $\eta_G^2 = .024$, that was qualified by an interaction of *material* and *generation task*, $F(2, 169) = 3.65$, $p = .028$, $\eta_G^2 = .023$: The effect of *material* was restricted to the free-generation condition,

$F(2, 83) = 4.84, p = .010, \eta_G^2 = .063$; it was not found in the cued-generation condition, $F(2, 86) = 1.70, p = .189, \eta_G^2 = .019$. We further analysed the effect of material in the free-generation condition using Tukey's HSDs: The *random* group generated more reversals than the other groups (*no-learning* group, $p = .012$, *permuted* group, $p = .053$), which did not differ from each other, $p = .897$.

Finally, and perhaps most importantly, as apparent from Figure 3 the proportion of reversals was below chance for all materials. This effect was most prominent in the *no-learning* group which had the smallest proportion of reversals. The overall below-chance generation proportions of reversals can therefore not be interpreted as an effect of training.

It is more likely that they reflect a response bias that participants bring into the lab (e.g., Boyer et al., 2005). If reversals represent a regular pattern, according to participants' subjective theory of randomness, they should tend to avoid generating such regularities when attempting to produce a random sequence. The finding that more reversals were generated under exclusion conditions than under inclusion conditions would then reflect participants' attempt to generate a non-random sequence under exclusion conditions that is most dissimilar to the random sequence from the training phase. This strategy may underlie the $I > E$ pattern of correct responses obtained above that would suggest the presence of explicit knowledge: In cases where reversals occurred at chance levels in the sequence, the tendency to avoid generating reversals would lead to an underestimation of (implicit and/or explicit) knowledge. In the present case, in which no reversals were encountered during learning, suppressing reversals leads to overestimated implicit knowledge. If this tendency is stronger under inclusion instructions than under exclusion instructions, as we have found here, it may erroneously suggest that participants have acquired explicit knowledge. Thus, the proportion of reversals that differed across PD instructions may be responsible for the above findings.

Correct generation performance when excluding reversals. Reversal generation proportions, whether reflective of extra-experimental response biases or not, may have distorted the above PD analyses and erroneously suggested the presence of implicit

and/or explicit knowledge. We therefore repeated the above analyses after excluding reversals, to investigate whether the above patterns of $I > E$ and $E > B$ hold also for the remaining types of transitions. Figure 4 shows the proportion of correct generation responses after removal of reversals.

After excluding reversals from analyses, an ANOVA revealed only a significant interaction of *material* \times *generation task*, $F(2, 169) = 4.39$, $p = .014$, $\eta_G^2 = .031$, and a trend towards a *generation task* by *PD instruction* interaction, $F(1, 169) = 3.46$, $p = .065$, $\eta_G^2 = .008$. Crucially, the main effect of *PD instruction* was no longer significant, $F(1, 169) = 0.42$, $p = .519$, $\eta_G^2 = .001$.

As above, the effect of *material* was limited to the free-generation task, $F(2, 83) = 4.53$, $p = .014$, $\eta_G^2 = .053$, it was absent from the cued-generation task, $F(2, 86) = 1.24$, $p = .296$, $\eta_G^2 = .019$. Tukey's HSDs replicated the above finding that the *permuted* group generated more correct responses than the *random* group, $p = .001$, and tended to generate more correct responses than the *no-learning* group, $p = .083$. The *random* and *no-learning* groups did not differ from each other, $p = .257$.

Excluding reversals eliminated the main effect of PD instruction, an effect that is typically interpreted as evidence for explicit knowledge. This shows that conclusions drawn from applications of PD to the SRTT can be distorted by response-tendency artifacts: Participants appeared to possess explicit and controllable knowledge about the material but in fact merely avoided generating reversals, especially under inclusion instructions.

High-frequency locations. Next, we investigated effects of the response location frequency manipulation in the *permuted* condition. The above differences in correct generation performance between *permuted* and the other two groups might be explained by the fact that the *permuted* group was able to acquire knowledge about the unequally distributed location frequencies: The *permuted* group was trained on 8-response sequences constructed from six response locations, with two of the locations doubled. For a given participant, the two selected locations remained constant throughout the training phase and

were therefore practiced more frequently than the other four response locations. This was reflected in faster responses at high-frequency compared to low-frequency response locations (see above, Figure 1).

In the absence of learning, one would expect one third (i.e., two out of six) of generated responses to be high-frequency locations. If participants in the permuted condition acquired and used knowledge to generate more high-frequency locations, this would increase their chances of producing a correct triplet. This is because approximately two thirds of the possible triplets end in a high-frequency location (in the *permuted* learning material, on average, a high-frequency location was the correct response for approximately 22 out of 30 transitions). Thus, the probability of generating a correct triplet may have been inflated by a tendency to generate an above-chance proportion of high-frequency locations.

The proportion with which high-frequency locations were generated is illustrated in Figure 5. We analyzed these proportions by way of a 2 (*generation task*: free vs. cued) \times 2 (*PD instruction*: inclusion vs. exclusion) \times 2 (*order*: inclusion first vs. exclusion first) ANOVA. Consistent with the above finding that participants responded faster to high-frequency locations during training, we found evidence that participants learned to match the response frequencies during generation: responses encountered more frequently during the SRTT were also generated more frequently. Yet, this tendency was affected by the type of *generation task*, $F(1, 52) = 13.97$, $p < .001$, $\eta_G^2 = .144$: more high-frequency locations were generated in the free generation than in the cued generation test. Additionally, there was a trend towards an interaction of *generation task* and *PD instruction*, $F(1, 52) = 3.04$, $p = .087$, $\eta_G^2 = .021$.

For free generation, the proportion of high-frequency locations generated was above the 1/3 chance level in both the inclusion blocks, $t(26) = 2.95$, $p = .007$, $d = 0.57$, and in the exclusion blocks, $t(26) = 2.47$, $p = .020$, $d = 0.48$. The proportion was *below* 1/3 for cued generation in the inclusion blocks, $t(28) = -2.58$, $p = .015$, $d = -0.48$, but not in the exclusion blocks, $t(28) = -0.84$, $p = .407$, $d = -0.16$. The proportion of high-frequency

locations in cued generation may be influenced by cueing artifacts as illustrated below.

In free generation, there was a tendency to generate high-frequency response *locations* at above-chance levels. As suggested above, this tendency might explain the above-chance generation of correct *triplets* in the permuted condition that remained after excluding reversals.

Effect of elevated high-frequency location generation rates on correct generation performance. To provide support for this interpretation, we investigated whether an above-chance generation rate as observed in the permuted condition is necessary and sufficient to account for this pattern. First, we corrected for the trend to generate more high-frequency locations by equally weighing correct generations for each response location.

For the *permuted* group we calculated the proportions of correctly generated triplets (excluding repetitions and reversals) separately for triplets that were completed by a high-frequency location and those completed by a low-frequency location. We then calculated the weighted mean correct performance, with one third as a weight for the two high-frequency locations and two thirds as the weight for low-frequency locations. Analyses of the equally-weighted free-generation data no longer revealed any significant effects. Crucially, the effect of material was no longer significant, $F(2, 83) = 1.44$, $p = .243$, $\eta_G^2 = .017$. Perhaps trivially so, all significant findings disappeared after we eliminated the greater weight of the high-frequency locations in determining the proportion of correct triplets. This suggests that the higher proportion of high-frequency locations was necessary to produce the above effect.

To test whether the only slightly above-chance generation of high-frequency responses was sufficient to explain the above pattern of correctly generated triplets, we simulated data for each participant based only on their average proportions of high-frequency and low-frequency responses.⁵

For a participant's simulated dataset, we then determined the proportions of correctly generated triplets as above (i.e., based on the material in the learning phase, and after

excluding reversals). We conducted the same analysis as above on the simulated data for the free-generation task. Importantly, the simulated data replicated the effect of material, $F(2, 87) = 49.64$, $p < .001$, $\eta_G^2 = .447$: Tukey's HSDs revealed significant differences between *permuted* and the other two groups, both $ps < .001$, and no difference between *random* and *no-learning* group, $p = .986$.

The above results demonstrate that the slightly elevated generation rate of high-frequency responses was necessary and sufficient to produce the performance advantage in the permuted-material condition. They illustrate how sequence-unrelated frequency properties of the training material may affect generation performance, and thereby, PD estimates of underlying processes.

Effects of generation task

Whereas the proportion of reversals as well as high-frequency locations affected performance in free generation, similar influences were not observed on cued-generation data. This suggests that the cued generation task may be less sensitive to subtle effects of learning. The cued generation task has been criticized because the cues may contain information about the sequence that may affect generation performance and distort estimates of learning. Here we investigated potential effects of cues on generation responses that may occur even in control conditions and in the absence of informational influence.

As cues, participants were presented with brief segments of response locations taken from the learning phase. It is known that recent responses may be less likely to be generated (e.g., Boyer et al., 2005). This bias may affect generation performance selectively in the *permuted* condition. Because the permuted material contained high-frequency as well as low-frequency response locations, the same was true for the cues presented to participants in the permuted condition during the cued-generation task. As a consequence, the bias to avoid recent locations would apply more strongly to high-frequency locations. This could account for the above finding that high-frequency locations were generated at below-chance levels in

the cued generation task, and for the suppressed levels of correct performance in the permuted condition in that task.

For the cued generation task, Figure 6 shows the proportion of response locations that were also presented as a cue on their respective trial. With three to five cues presented on each trial (and direct repetitions prohibited and excluded from analyses), chance level of generating a location that had just been presented equalled $3/5 = .6$.

A 3 (*material*: permuted, random, no-learning) \times 2 (*PD instruction*: inclusion vs. exclusion) \times 2 (*block order*: inclusion first vs. exclusion first) ANOVA revealed a main effect of *material*, $F(2, 86) = 3.41$, $p = .038$, $\eta_G^2 = .042$. Tukey's HSDs revealed a significant difference between the *permuted* and *no-learning* groups, $p = .026$, all other $ps > .24$. The ANOVA also revealed a main effect of *PD instruction*, $F(1, 86) = 10.24$, $p = .002$, $\eta_G^2 = .051$: more repetitions of cue locations were generated under exclusion conditions. This effect was qualified by an interaction *PD instruction* \times *block order*, $F(1, 86) = 6.06$, $p = .016$, $\eta_G^2 = .031$, which indicates that the effect of PD instruction differed between participants who worked under inclusion instructions in the first generation block and under exclusion instructions in the second block and those who received the PD instructions in the reverse order. To further explore this interaction, we analysed first and second blocks using separate ANOVAs, thus turning the *PD instruction* factor into a between-subjects factor. In the first block, only the main effect of *material* turned out to be significant, $F(2, 86) = 3.67$, $p = .030$, $\eta_G^2 = .079$. In the second block, only the main effect of *PD instruction* was significant, $F(1, 86) = 10.91$, $p = .001$, $\eta_G^2 = .113$.⁶

These results suggest three conclusions: First, the fact that the tendency to avoid generating cued locations was also present in the no-learning condition suggests that it was not acquired during the SRTT training but either reflects pre-experimental response tendencies, or, alternatively, it may reflect a tendency of the cued generation task to bias generation performance against repeating the cue locations.

Second, the cue-avoidance bias was influenced by the type of learning material in the

first but not the second block. This suggests that, while the cued generation task is sensitive to effects of learning during the first block, this is no longer the case during the second block. This finding further supports the notion that the cued generation format may affect generation performance, and that over time this influence becomes stronger than the effects of learning.

Third, cue-avoidance bias was stronger under inclusion than under exclusion instructions in the second block of the generation task. This suggests that participants can acquire response strategies during the cued-generation task that help them produce different outcomes under inclusion and exclusion instructions.

Whether or not this tendency leads to artificially elevated levels of correct responses largely depends on the sequence material that is used. Common sequences (e.g., Q. Fu et al., 2010, 2008; Wilkinson & Shanks, 2004) contain only one reversal within a four-position sequence consisting of 12 triplets. Typically, only two locations are presented as cues on each trial to avoid informative influences, and direct repetitions are not allowed. In this case, the strategy to avoid generating a cue location would lead to a correct-performance rate of approximately 46% (i.e., a .5 chance to generate a correct response for 11 out of 12 triplets and a zero chance for the single reversal among the 12 triplets). This is rather high relative to a chance-level baseline of .33 (c.f., Q. Fu et al., 2010).

To summarise, in the cued generation task, participants avoid generating a response that had been part of the cue sequence. This finding was independent of the type of training material, and it occurred even for the no-learning group, suggesting that it is a training-independent response tendency (i.e., a bias against generating locations present in the cue segment).

The findings further suggest that, in the permuted condition, this response strategy interacted with cue properties to influence generation performance. Here, cue properties vary as a function of the frequency manipulation in the permuted condition: Frequent locations are more likely to be included in the cue, and are therefore more likely to be subject to the

avoidance bias; this can explain the reduction, in the *permuted* condition, of high-frequency responses in the cued as compared to the free generation task.

Finally, the cue-avoidance bias differed across PD instructions: cued responses were generated less frequently under inclusion than under exclusion instructions. Interestingly, this difference emerged only in the second generation block, suggesting that the cued generation task allowed participants to acquire a strategy of selectively generating more cue locations under exclusion instructions: If participants perceived the learning material as random, they will aim at generating a random sequence under inclusion instructions (i.e., generate a sequence similar to that in the learning phase); under exclusion instructions, they will aim at generating a sequence that does not conform to their subjective notion of randomness (i.e., generate a sequence that is as dissimilar as possible to that in the learning phase). When generating subjectively random sequences, participants typically produce more alternations and fewer repetitions than would be expected by chance (e.g., Boyer et al., 2005). By attempting to deviate from this subjectively experienced randomness they then generate more cue locations.

Taken together, the cued generation task was not only less sensitive to learning effects. It also appears to induce – or interact unfavorably with – a bias to avoid recently generated response locations. Both the below-chance generation of cue locations as well as its modulation by PD instruction may distort findings and conclusions regarding the presence or absence of explicit and implicit sequence knowledge.

Discussion

The present findings extend previous knowledge about the influence of pre-experimental response biases, simple frequency information, task format, and their interaction on performance in the generation task. First, it is known that zero-order frequency information may affect generation performance (Reed & Johnson, 1994); here, we show that this effect depends on the type of generation task. Second, it is known that, in

cued generation, cues that carry information about the sequence may affect generation performance (Q. Fu et al., 2010). We present first evidence that even in the absence of any sequence information, cues may affect generation performance by way of their simple frequency properties and in interaction with response tendencies. Third, it is known that participants may be biased to avoid generating recent response locations (Boyer et al., 2005). Here, we demonstrated that this bias may interact with properties of the learning material (zero-order frequencies, proportion of reversals) and task format, and that it can differentially affect inclusion and exclusion performance. These findings have important implications for the validity of the process-dissociation procedure as applied to the generation task.

Process-dissociation results

The present study realized three different control conditions, using different types of pseudo-random materials, in which participants could not learn any second-order regularity. Implicit and explicit knowledge was assessed with the PD procedure, using two different generation tasks. Despite the irregular learning materials, and independent of the type of analysis, the results of a comparison between performance in the permuted and random conditions in the free-generation task suggested the presence of implicit knowledge. Similarly, independent of the type of analysis, the results also suggested the presence of explicit knowledge under some conditions (i.e., due to the asymmetric generation of reversals under inclusion versus exclusion instructions). Finally, the results erroneously underestimated knowledge in some conditions: Although participants learned to distinguish between high- and low-frequency locations, they expressed this knowledge in the free-generation condition but did not express this knowledge in the cued-generation condition. These findings illustrate that unwanted influences can affect generation performance and may thereby artificially inflate or mask estimates of implicit and explicit knowledge.

Taken together, the findings suggest that there are several problems with the PD approach in its application to investigating the processes underlying sequence learning when

effects of extra-experimental influences on performance cannot be excluded. In those cases, baseline performance may differ in inclusion and exclusion instructions and/or across experimental conditions. Under those circumstances, the ordinal PD approach is no longer valid (Hirshman, 2004).⁷ In other words, given the possible influence of response tendencies and their interaction with properties of the material, we cannot draw conclusions about the relative contribution of implicit and explicit knowledge by simply comparing inclusion and exclusion performance, be it directly (e.g., $I > E$), across experimental conditions (e.g., $E_A > E_B$), or with a baseline (e.g., $E > B$). Instead, it is necessary to quantify the unwanted extraneous influence and separate it from the experimental effects of interest. This can be done by extending the basic PD design and model.

An extended PD model

In applications of the PD, exclusion performance is sometimes compared to an a-priori chance level baseline. The present study has shown that the proportion of correctly generated triplets may deviate from such an a-priori chance-level baseline, and may even differ across inclusion and exclusion conditions, in the absence of both implicit and explicit sequence knowledge. In other words, we found that response tendencies, alone or in interaction with properties of the learning material, may affect generation performance, may cause deviations from a-priori chance baselines, and may distort the PD model's estimates of implicit and explicit knowledge.

To accommodate this potential confound, first, it is necessary to use an empirical baseline. One could extend the PD design and model, as applied to the generation task, by adding a control condition along with separate response-tendency or nuisance parameters, as has been done by Buchner and colleagues for the recollection task (Buchner et al., 1998, 1997). In the control condition, that is, in the absence of sequence knowledge, generation performance will then reflect sequence-unspecific properties of the material as well as response tendencies. In the experimental condition, these same influences may affect

generation performance over and above the effects of learning higher-order sequence information. The nuisance parameter can then be equated across the control and experimental conditions, capture the unwanted influences, and separate them from the effects of implicit and explicit knowledge. In this model, any differences in performance between the control and experimental conditions will be reflected in estimates of implicit and/or explicit knowledge. Second, the problem that response tendencies may differentially affect inclusion and exclusion performance can be solved by using as separate baselines the performance of a control group in the inclusion and exclusion conditions, and by introducing one nuisance parameter for each condition. Such an extended model allows for the possibility that response tendencies affect generation performance over and above explicit and implicit knowledge. Similar extensions for capturing extraneous influences have been proposed and demonstrated as necessary and useful in other domains to address similar response-tendency issues (e.g., pre-experimental familiarity, Buchner et al., 1995; right-hand bias; Stahl & Degner, 2007).

The extended model's equations deviate from the original equations (see introduction) by including nuisance parameters that account for baseline performance (along the lines suggested by Buchner et al., 1995, 1997). Because baseline performance may be different under inclusion and exclusion instructions, two different parameters $R_{inclusion}$ and $R_{exclusion}$ are added to the model. The equations of the experimental conditions of the extended model are: $p(\text{correct}|\text{inclusion}, \text{experimental}) = C + (1 - C) * A + (1 - C)(1 - A) * R_{inclusion}$, and $p(\text{correct}|\text{exclusion}, \text{experimental}) = (1 - C) * A + (1 - C)(1 - A) * R_{exclusion}$. For the empirical control or baseline condition, it would be assumed that only extra-experimental or nuisance factors affect generation performance: $p(\text{correct}|\text{inclusion}, \text{control}) = R_{inclusion}$, and $p(\text{correct}|\text{exclusion}, \text{control}) = R_{exclusion}$.

To illustrate, we applied this model to the permuted and random conditions in the free generation task. Recall that both traditional approaches suggested the presence of explicit and implicit knowledge in the permuted group, as well as the presence of explicit knowledge

in the random group. The extended model allows us to quantify the effect of training with permuted material over a training phase with random material on controlled and automatic processes. We used the random group as the empirical control condition; this is equivalent to the assumption that whatever affects performance in the random group can be subsumed as nuisance factors. Estimates of $R_{inclusion}$ and $R_{exclusion}$ therefore reflect generation performance in the random group, $R_{inclusion} = .251$ and $R_{exclusion} = .228$. Differences between the random and permuted condition are then reflected in the parameters for controlled and automatic processes. Results of the extended model suggests that training with permuted material affected the automatic process, $A = .047$, $\Delta G^2_{(df=1)} = 16.44$, $p < .001$, but did not affect the controlled process, $C = .001$, $\Delta G^2_{(df=1)} = 0.004$, $p = .95$. The effect on the automatic process – implicit knowledge – reflects the location frequency effect, and it is consistent with the finding that participants responded faster to high-frequency than to low-frequency locations during SRTT training. The extended model indicates the absence of an effect on explicit knowledge. The nuisance parameters capture the difference between inclusion and exclusion performance in both the random and permuted group that was interpreted as evidence for explicit knowledge in both the ordinal PD approach as well as in the original model equations.

Note that the levels of generation performance in the control condition can serve as valid baselines – and the controlled and automatic parameters can be valid estimates – only to the degree that the processes captured by the nuisance parameters are independent from explicit and implicit knowledge processes. In other words, the extended model requires additional assumptions, namely that the processes underlying the parameters C and A are independent from the nuisance processes (e.g., guessing, response bias) that are reflected in the new parameters $R_{inclusion}$ and $R_{exclusion}$. These assumptions are as of yet untested, and future research is needed to supply empirical evidence as to whether these new independence assumptions are met.

Free versus cued generation

In free generation, participants sometimes produce relatively sparse data, for instance, by repeatedly generating a small subset of all possible transitions, thereby generating missing data for the remaining transitions. With the cued generation task, researchers can control the cues and thereby ensure that comparable (or at least considerable) numbers of observations are obtained for each transition. This is an advantage especially when different types of transitions and their properties are to be compared in a within-subject approach.

However, the present findings support previous warnings that the cued generation task must be treated with caution because the choice of cues may influence generation performance (Q. Fu et al., 2010; Jiménez & Vázquez, 2005). In addition to previous findings regarding informative cues, the present study provides an additional argument against the cued generation task: The tendency to avoid generating locations that were presented as cues may systematically bias generation performance, and may erroneously suggest the presence of both implicit and explicit knowledge.

In light of these problems, we currently do not see an empirical way of obtaining an equal number of observations for each transition in the generation task. This could imply that attempts of modeling individual transitions as a within-participant factor will face additional challenges. One way of avoiding artifacts resulting from a limited and biased selection of generated transitions is to use adequate weighting procedures. For instance, researchers could compute correct-generation rates for each of the cells in the transition matrix and estimate participants' mean proportion of correctly generated transitions as a weighted average.

Limitations and open questions

We briefly sketch open research questions posed by the additional independence assumptions, discuss the current development of hierarchical multinomial modeling techniques, and note the potential influence of additional unknown factors affecting

performance in the generation task.

On the independence assumptions underlying the extended model. As mentioned above, the extended PD model assumes that response tendencies as modelled by $R_{inclusion}$ and $R_{exclusion}$ are independent of controlled and automatic processes. If this assumption is violated, parameter estimates are no longer valid measures of the underlying psychological processes. This possibility deserves to be taken seriously; in other domains, empirical violations of independence between controlled and automatic processes have been obtained (e.g., Rouder et al., 2008). The open question is whether response tendencies can be affected by the same properties of the learning material that may also affect implicit and explicit learning. There are some findings in the literature that at least suggest such an interaction may well be possible. For instance, participants tended to show lag effects (reflecting negative recency) in a permuted condition but not in a random condition (Boyer et al., 2005). Here, the tendency to avoid generating recent locations interacted with properties of the learning material. Perhaps more critically, the present results suggest that in the permuted condition, the bias against generating cued locations was reduced whereas implicit knowledge was enhanced (i.e., participants learned about the frequency with which certain response locations occurred during training). If the response bias is reduced in a condition with greater implicit knowledge, then the response bias estimates (i.e., the R parameters) obtained from the control condition (where implicit knowledge is lower or entirely absent) will tend to overestimate the response bias in that condition. The present finding may reflect effects of other factors, but if such an interaction pattern between response tendency and implicit knowledge can be substantiated, it would constitute evidence against the independence assumption. In turn, this would require further elaborating and refining the extended PD paradigm and model.

Data aggregation and hierarchical modeling. In our application of the PD model, we aggregated the data across participants, as this has been the standard procedure in common applications of PD and our aim was to illustrate potential problems with such

applications. Our results demonstrate that the PD procedure, as commonly applied, yields erroneous conclusions supporting the presence of implicit and/or explicit knowledge, even in the absence of such knowledge. Note that such erroneous conclusions about the presence of an influence on parameters are less likely when, instead of aggregating, individual data are analyzed using hierarchical models. This is because data aggregation assumes parameter homogeneity across participants, which is likely to be violated in most cases. As a consequence of aggregation in the presence of heterogeneity, confidence intervals can be underestimated, and nominal statistical error levels can be violated by statistical tests.⁸

This problem does not arise if hierarchical modeling techniques are used (Klauer, 2006, 2010; e.g., Rouder et al., 2008) that account for the heterogeneity across participants (and items) in their estimates of the parameters' variability. Hierarchical models often draw a more realistic picture of the variability of parameter estimates. In all cases in which parameters must be assumed to vary across participants, they should be preferred to the traditional approach of data aggregation.⁹

Characteristics of response tendencies. In the present study, we obtained evidence that participants may use two types of response tendencies: First, they generated fewer reversals than expected by chance. Second, in the cued generation task, they avoided generating responses that were part of the cue segments. It is unclear whether these patterns reflect the same bias (e.g., based on subjective notions of randomness) or whether they reflect different response tendencies, and whether they reflect consciously applied strategies or rather more implicit trends. In addition, results obtained with the extended model suggested that, even in the absence of both implicit and explicit knowledge and after controlling for the identified response tendencies, there are other unknown factors that allow participants to perform better than chance, and even better under inclusion than under exclusion conditions. Future research should address these questions in order to better understand their potential for affecting generation performance.

Conclusion

Broadly speaking, the process-dissociation approach is to create two (or more) experimental conditions within a given experimental paradigm so that there is overlap with regard to most but not all of the psychological processes that are relevant for performance. Combined with measurement models such as the PD equations or more complex multinomial models, this general approach has been extremely valuable and has been fruitful in a wide variety of research areas and experimental paradigms (Erdfelder et al., 2009; Yonelinas & Jacoby, 2012). Whereas a simple and elegant design is generally preferable, the present findings suggest that the application of process-dissociation methodology to the generation task may require more differentiation of experimental design and measurement model.

References

- Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., . . . Hyndman, R. (2015). *rmarkdown: Dynamic Documents for R*. Retrieved from <http://CRAN.R-project.org/package=rmarkdown>
- Aust, F., & Barth, M. (2015). *papaja: Create APA manuscripts with RMarkdown*. Retrieved from <https://github.com/crsh/papaja>
- Boyer, M., Destrebecqz, A., & Cleeremans, A. (2005). Processing abstract sequence structure: Learning without knowing, or knowing without learning? *Psychological Research*, *69*(5-6), 383–398. doi:[10.1007/s00426-004-0207-4](https://doi.org/10.1007/s00426-004-0207-4)
- Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, *124*(2), 137–160. doi:[10.1037/0096-3445.124.2.137](https://doi.org/10.1037/0096-3445.124.2.137)
- Buchner, A., Steffens, M. C., & Rothkegel, R. (1998). On the role of fragmentary knowledge in a sequence learning task. *The Quarterly Journal of Experimental Psychology Section A*, *51*(2), 251–281. doi:[10.1080/713755757](https://doi.org/10.1080/713755757)
- Buchner, A., Steffens, M. C., Rothkegel, R., & Erdfelder, E. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 631–663. doi:[10.1080/713755723](https://doi.org/10.1080/713755723)
- Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 531–547. doi:[10.1037/0278-7393.21.3.531](https://doi.org/10.1037/0278-7393.21.3.531)
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*(2), 343–350. doi:[10.3758/BF03196171](https://doi.org/10.3758/BF03196171)
- Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task.

Consciousness and Cognition, 19(2), 674–681. doi:[10.1016/j.concog.2009.09.009](https://doi.org/10.1016/j.concog.2009.09.009)

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Abfal, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift Für Psychologie / Journal of Psychology*, 217(3), 108–124. doi:[10.1027/0044-3409.217.3.108](https://doi.org/10.1027/0044-3409.217.3.108)

Fu, Q., Dienes, Z., & Fu, X. (2010). Can unconscious knowledge allow control in sequence learning? *Consciousness and Cognition*, 19(1), 462–474.

doi:[10.1016/j.concog.2009.10.001](https://doi.org/10.1016/j.concog.2009.10.001)

Fu, Q., Fu, X., & Dienes, Z. (2008). Implicit sequence learning and conscious awareness. *Consciousness and Cognition*, 17(1), 185–202. doi:[10.1016/j.concog.2007.01.007](https://doi.org/10.1016/j.concog.2007.01.007)

Haider, H., Eichler, A., & Lange, T. (2011). An old problem: How can we distinguish between conscious and unconscious knowledge acquired in an implicit learning task?

Consciousness and Cognition, 20(3), 658–672. doi:[10.1016/j.concog.2010.10.021](https://doi.org/10.1016/j.concog.2010.10.021)

Hirshman, E. (1998). On the logic of testing the independence assumption in the process-dissociation procedure. *Memory & Cognition*, 26(5), 857–859.

doi:[10.3758/BF03201168](https://doi.org/10.3758/BF03201168)

Hirshman, E. (2004). Ordinal Process Dissociation and the Measurement of Automatic and Controlled Processes. *Psychological Review*, 111(2), 553–560.

doi:[10.1037/0033-295X.111.2.553](https://doi.org/10.1037/0033-295X.111.2.553)

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541.

doi:[10.1016/0749-596X\(91\)90025-F](https://doi.org/10.1016/0749-596X(91)90025-F)

Jiménez, L., & Vázquez, G. A. (2005). Sequence learning under dual-task conditions: alternatives to a resource-based account. *Psychological Research*, 69(5-6), 352–368.

doi:[10.1007/s00426-004-0210-9](https://doi.org/10.1007/s00426-004-0210-9)

Klauer, K. C. (2006). Hierarchical Multinomial Processing Tree Models: A Latent-Class Approach. *Psychometrika*, 71(1), 7–31. doi:[10.1007/s11336-004-1188-3](https://doi.org/10.1007/s11336-004-1188-3)

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A

Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98.

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2013). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. Retrieved from <http://link.springer.com/article/10.1007/s11336-013-9374-9>

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. doi:[10.1016/0010-0285\(87\)90002-8](https://doi.org/10.1016/0010-0285(87)90002-8)

Norman, E., Price, M. C., & Duff, S. C. (2006). Fringe consciousness in sequence learning: The influence of individual differences. *Consciousness and Cognition*, *15*(4), 723–760. doi:[10.1016/j.concog.2005.06.003](https://doi.org/10.1016/j.concog.2005.06.003)

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261. doi:[10.1038/nn1840](https://doi.org/10.1038/nn1840)

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 585–594. doi:[10.1037/0278-7393.20.3.585](https://doi.org/10.1037/0278-7393.20.3.585)

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, *137*(2), 370–389. doi:[10.1037/0096-3445.137.2.370](https://doi.org/10.1037/0096-3445.137.2.370)

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560–575. doi:[10.3758/s13428-012-0259-0](https://doi.org/10.3758/s13428-012-0259-0)

Sinnwell, J. P., & Schaid, D. J. (2013). *ibdreg: Regression Methods for IBD Linkage With Covariates*. Retrieved from <http://CRAN.R-project.org/package=ibdreg>

Stadler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 18(2), 318–327.

doi:[10.1037/0278-7393.18.2.318](https://doi.org/10.1037/0278-7393.18.2.318)

Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology*, 54(2), 99–112.

doi:[10.1027/1618-3169.54.2.99](https://doi.org/10.1027/1618-3169.54.2.99)

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, 39(2), 267–273.

doi:[10.3758/BF03193157](https://doi.org/10.3758/BF03193157)

Wilkinson, L., & Shanks, D. R. (2004). Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 354–369.

doi:[10.1037/0278-7393.30.2.354](https://doi.org/10.1037/0278-7393.30.2.354)

Xie, Y. (2013). *Dynamic Documents with R and knitr*. Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from <http://yihui.name/knitr/>

Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition*, 40(5), 663–680. doi:[10.3758/s13421-012-0205-5](https://doi.org/10.3758/s13421-012-0205-5)

Footnotes

¹ Note that the $E > B$ pattern could not be replicated in other studies (Norman, Price, & Duff, 2006; Wilkinson & Shanks, 2004) that, as a baseline, compared generation performance with that for a *control sequence* instead of an a-priori fixed value.

² For all our analyses, we used R (3.2.1, R Core Team, 2015) and the R-packages *ibdreg* (0.2.5, Sinnwell & Schaid, 2013), *knitr* (1.10.5, Xie, 2013), *papaja* (0.1.0.9054, Aust & Barth, 2015), and *rmarkdown* (0.7, Allaire et al., 2015).

³ Because we only used pseudo-random sequences in this study, we defined that, given the last two locations, the location that was most frequently presented after these two locations during training as being the ‘correct’ response. For the *no-learning* condition, correct responses were computed on the basis of sequences identical to those used in the *random* group.

⁴ Restricting parameters to be equal across *generation task* (free vs. cued) harmed goodness of fit, $\Delta G^2_{(df=6)} = 68.14, p < .001$; therefore, the amount of controlled and/or automatic processes must be assumed to differ across tasks. Similarly, parameters could not be equated across materials (no-learning, permuted, random), $\Delta G^2_{(df=8)} = 81.59, p < .001$, implying effects of material.

⁵ All other information from their empirical distribution of response location frequencies as well as any order information that may be present in participants’ generation responses was discarded: A high-frequency (low-frequency) response location was simulated with the probability given by the participant’s rate of actually generating a high-frequency (low-frequency) response in the generation task; within the set of high-frequency (low-frequency) locations, each was generated with equal probability. Datasets were simulated by drawing responses from this distribution (with replacement, but with the constraints that no direct repetitions were allowed).

⁶ This data pattern remained even after excluding reversals: Analysing only the first block, only the main effect of *material* trended to be significant, $F(2, 86) = 2.70, p = .073$,

$\eta_G^2 = .059$. Analysing only the second block revealed a main effect of *PD instruction*, $F(1, 86) = 6.63$, $p = .012$, $\eta_G^2 = .072$.

⁷ The ordinal PD approach assumes that baseline performance does not differ across conditions or tasks (see Hirshman, 2004, p.559, section ‘Comparison of the Current Approach and the Process-Dissociation Procedure’ and Footnote 8). Such baseline differences may arise from different guessing or response bias, and they invalidate the conclusions that can be drawn from the comparison of performance in conditions. This is because the performance differences between conditions or tasks that are used to draw conclusions about differences in automatic and/or controlled processes may instead reflect baseline differences due to guessing or response bias.

⁸ We applied the hierarchical Bayesian PD model proposed by Rouder et al. (2008) to the present data in order to account for person and item variability. The resulting parameter estimates are available in the supplemental material that can be obtained from <https://github.com/methexp/pdl1>. The results corroborated the findings obtained with the traditional analyses reported above (i.e., $C > 0$, $A > .2$, and the ordering of A estimates across conditions).

⁹ The use and adoption of hierarchical modeling approaches is currently limited by the availability of general-purpose software (but see, e.g., Matzke, Dolan, Batchelder, & Wagenmakers, 2013; Stahl & Klauer, 2007).

Appendix A

Proportions of correctly generated triplets

Material	Generation task	Order	Inclusion		Exclusion	
			M	95% CI	M	95% CI
No-learning	Free	Inclusion first	.23	[.20, .27]	.21	[.18, .25]
No-learning	Free	Exclusion first	.22	[.17, .27]	.24	[.19, .30]
No-learning	Cued	Inclusion first	.22	[.17, .27]	.22	[.17, .28]
No-learning	Cued	Exclusion first	.28	[.22, .34]	.25	[.21, .28]
Permuted	Free	Inclusion first	.29	[.26, .32]	.23	[.18, .28]
Permuted	Free	Exclusion first	.27	[.22, .31]	.25	[.20, .29]
Permuted	Cued	Inclusion first	.24	[.20, .28]	.21	[.17, .25]
Permuted	Cued	Exclusion first	.19	[.13, .25]	.20	[.16, .23]
Random	Free	Inclusion first	.23	[.20, .25]	.20	[.15, .26]
Random	Free	Exclusion first	.21	[.18, .25]	.18	[.15, .21]
Random	Cued	Inclusion first	.23	[.19, .28]	.24	[.18, .29]
Random	Cued	Exclusion first	.23	[.20, .26]	.23	[.19, .26]

Note. CI = Confidence intervals

Appendix B

Estimates of model parameters before or after reversals were removed

Material	Generation task	Full dataset		After excluding reversals	
		A	C	A	C
No-learning	Free	.23 [.22, .24]	.00 [-.02, .02]	.24 [.23, .26]	.00 [-.02, .02]
No-learning	Cued	.24 [.23, .25]	.02 [.00, .04]	.27 [.26, .28]	.00 [-.02, .02]
Permuted	Free	.25 [.24, .26]	.04 [.02, .06]	.27 [.26, .28]	.03 [.01, .06]
Permuted	Cued	.21 [.20, .22]	.01 [-.01, .03]	.24 [.23, .25]	.00 [-.02, .02]
Random	Free	.20 [.19, .21]	.03 [.01, .05]	.23 [.22, .24]	.02 [.00, .05]
Random	Cued	.23 [.22, .24]	.00 [-.02, .02]	.26 [.25, .27]	.00 [-.02, .02]

Note. 95% confidence intervals are in parentheses.

Appendix C

Category counts

Material	Generation task	Inclusion		Exclusion	
		Correct	Incorrect	Correct	Incorrect
<i>Full dataset</i>					
No-learning	Free	851	2894	851	2875
No-learning	Cued	886	2626	825	2675
Permuted	Free	881	2285	755	2427
Permuted	Cued	723	2584	673	2621
Random	Free	777	2731	669	2824
Random	Cued	879	2916	888	2914
<i>After excluding reversals</i>					
No-learning	Free	851	2671	851	2605
No-learning	Cued	886	2452	825	2233
Permuted	Free	881	2101	755	2127
Permuted	Cued	723	2306	673	2112
Random	Free	777	2360	669	2324
Random	Cued	879	2642	888	2434

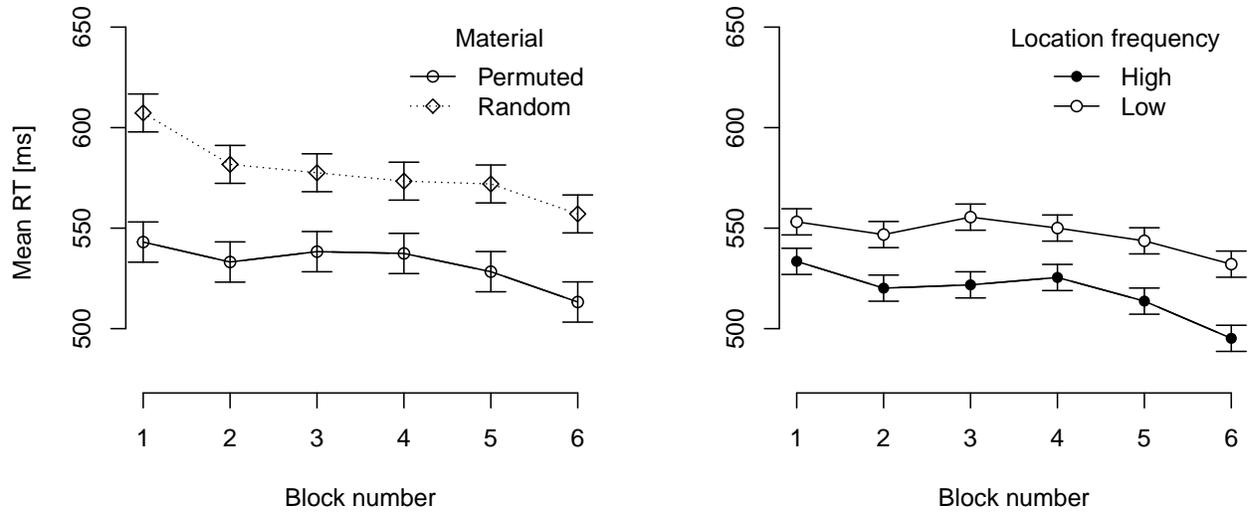


Figure 1. Left: Mean reaction times for permuted (solid line) and random material (dashed line). Right: Mean reaction times for permuted material, split by high-frequency (filled circles) vs. low-frequency locations (open circles). Error bars represent 95% within-subjects CIs.

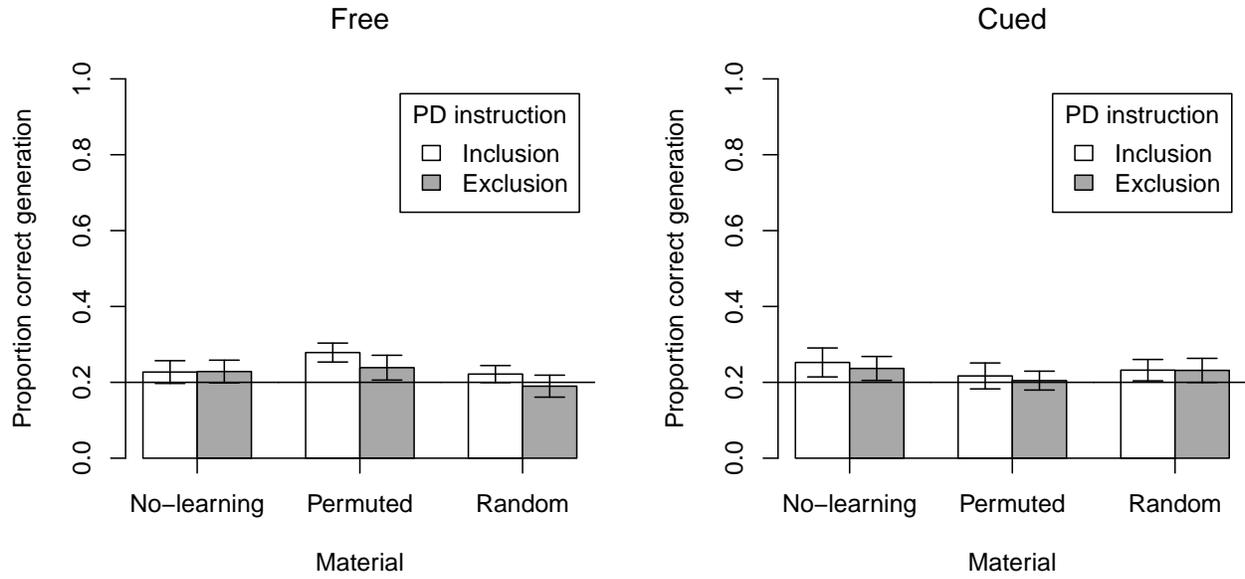


Figure 2. Proportions of correctly generated triplets during PD generation task (chance level = .2). Error bars represent 95% CIs.

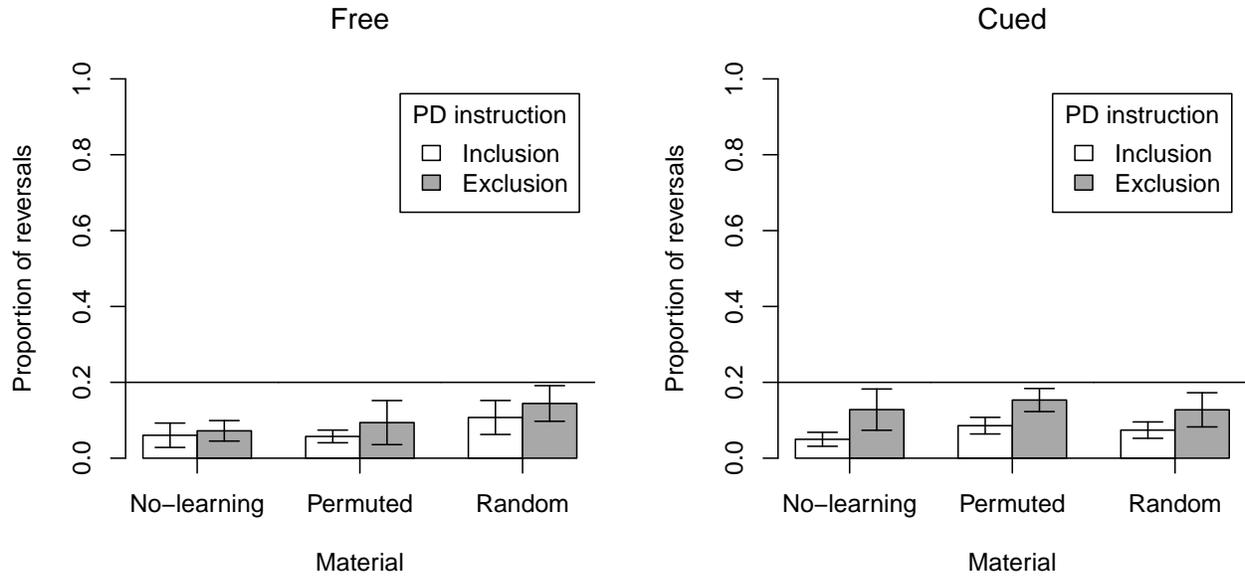


Figure 3. Proportions of reversals produced in generation task (chance level = .2). Error bars represent 95% CIs.

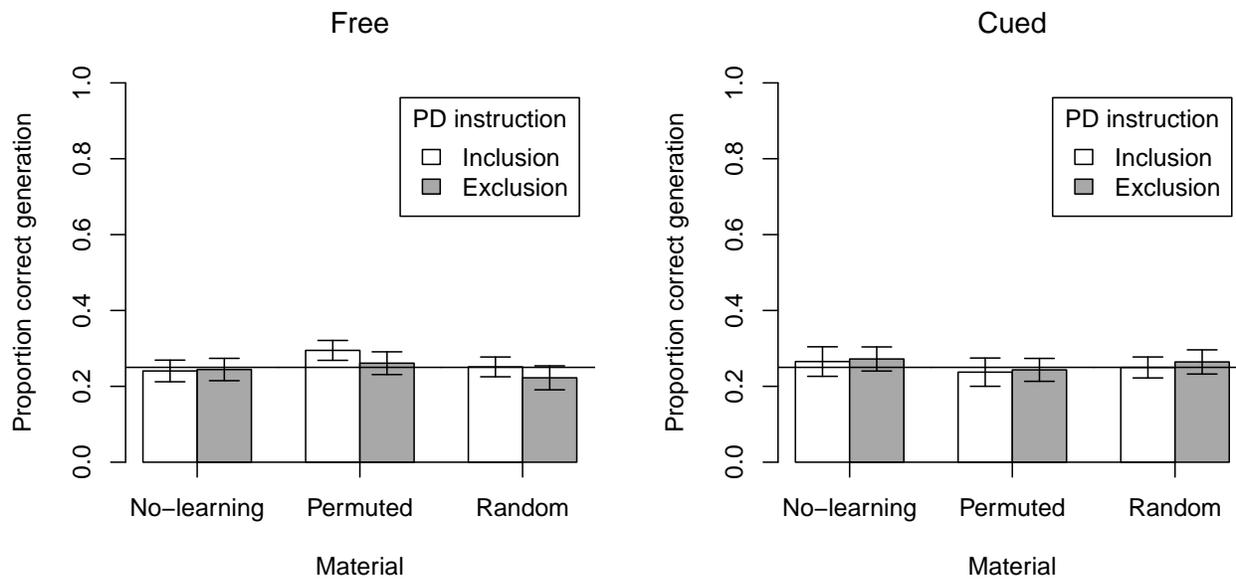


Figure 4. Proportions of correctly generated triplets during PD generation task after excluding reversals (chance level = .25). Error bars represent 95% CIs.

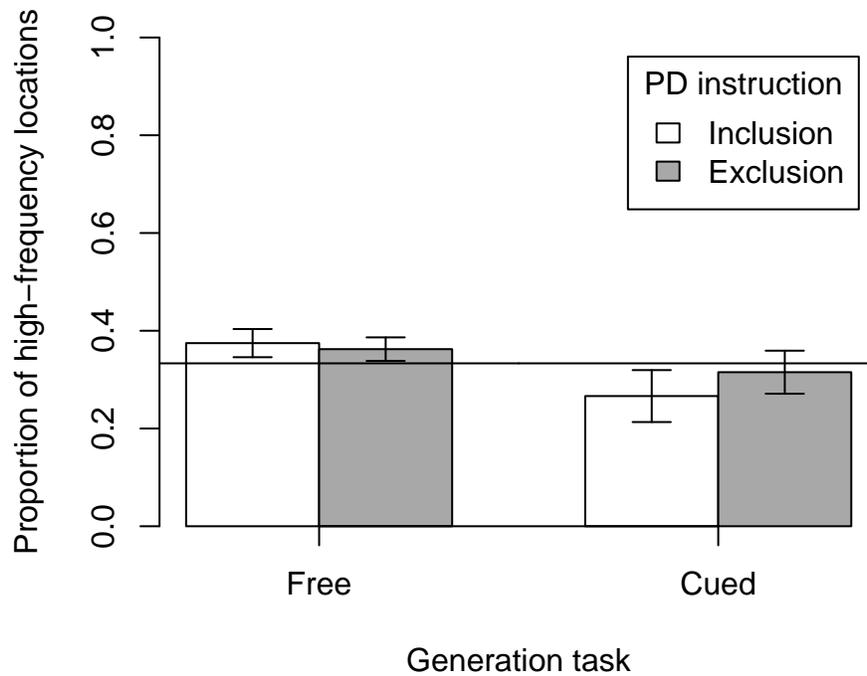


Figure 5. Proportion of high-frequency locations during generation in the permuted material condition (chance level = .33). Error bars represent 95% CIs.

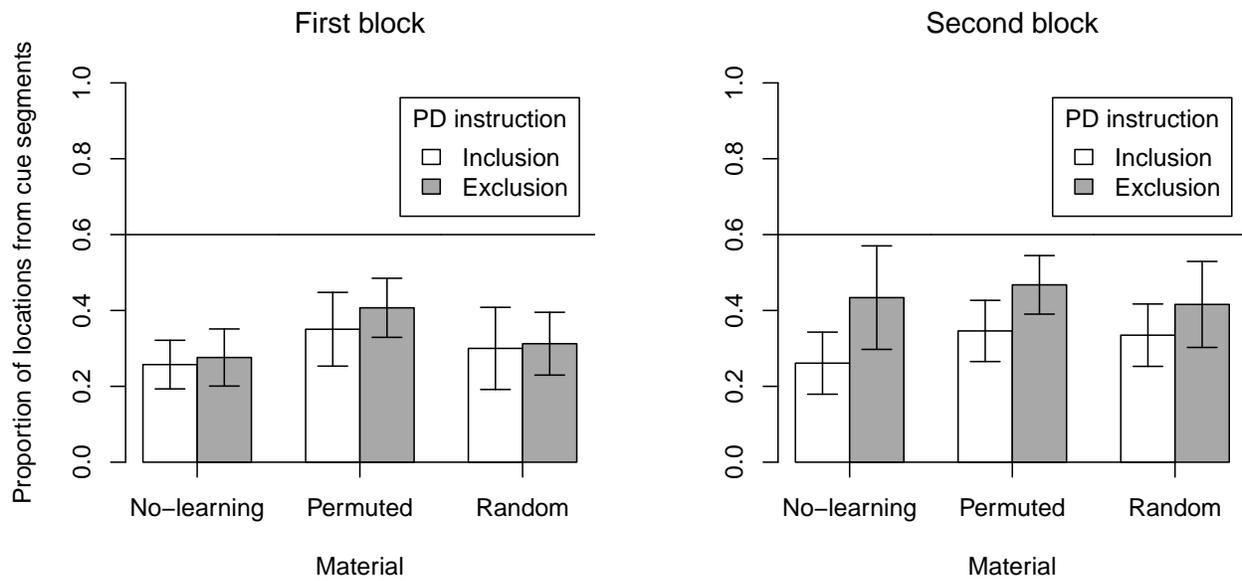


Figure 6. Proportions of locations generated in the cued generation task that were presented as a cue on their respective trial, separately for the first and the second generation block (chance level = .6). Error bars represent 95% CIs.