Transfer of strategic task components across unique tasks that share some common structures

Cai S. Longman[1,2], Fraser Milton[2], and Andy J. Wills[3]

[1] University of the West of Scotland, Paisley, UK

[2] University of Exeter, Exeter, UK

[3] University of Plymouth, Plymouth, UK

Abstract: 229 words. Text: 12424 words, 47 references, 4 figures, 7 tables.

Running head: Transfer of strategic task components

Corresponding author:
Cai S. Longman
Psychology
School of Education and Social Sciences
University of the West of Scotland
High Street
Paisley
PA1 2BE
United Kingdom
e-mail: cai.longman@uws.ac.uk
tel: +44 (0)141 8483039

Abstract

Flexible, adaptive behaviour depends on the application of prior learning to novel contexts (transfer). Transfer can take many forms, but the focus of the present study was on 'task schemas' – learning strategies that guide the earliest stages of engaging in a novel task. The central aim was to examine the architecture of task schemas and determine whether strategic task components can expedite learning novel tasks that share some structural components with the training tasks. Groups of participants across two experiments were exposed to different training regimes centred around multiple unique tasks that shared some/all/none of the structural task components (the kinds of stimuli, classifications, and/or responses) but none of the surface features (the specific stimuli, classifications, and/or responses) with the test task (a dot-pattern classification task). Initial test performance was improved (to a degree) in all groups relative to a control group whose training did not include any of the structural components relevant to the test task. The strongest evidence of transfer was found in the *motoric, perceptual + categorization,* and *full schema* training groups. This observation indicates that training with some (or all) strategic task components expedited learning of a novel task that shared those components. That is, task schemas were found to be componential and were able to expedite learning a novel task where similar (learning) strategies could be applied to specific elements of the test task.

Keywords: learning, transfer, abstract representation, task schema, categorization

Flexible, adaptive behaviour depends on the application of prior learning to novel contexts (transfer). Without the ability to transfer learned material beyond the learning context, any slight variation would require that a given task be learned 'from scratch'. Despite its importance in flexible adaptive behaviour and the education sector, the mechanisms by which transfer takes place (especially in the early stages of learning) remain poorly understood. The aim of the present study was to develop our understanding of transfer by investigating how the early stages of learning a novel task are guided by recently used learning strategies.

According to Chein and Schneider's (2012) 'triarchic' theory of learning, there are three distinct stages of learning (formation of a learning strategy, controlled execution, automatic execution), each of which is localized in a distinct neural network/system (metacognitive system, cognitive control network, representation system). The early stages of learning are characterized by the formation of a learning strategy or 'task schema', by the metacognitive system, which is then applied to the current context by the cognitive control network. Eventually, it becomes possible for the representation system to perform the (now well-practiced) task with little or no intervention from the other systems and performance is thought to be 'automatic'.

Application of a 'task-set' (a tuning of the cognitive control network toward completion of a specific task/goal; e.g., see Kiesel et al., 2010; Monsell, 2003; Vandierendonck, et al., 2010) will eventually result in the phased transition to automatic execution of the task. Similarly, early application of a task schema will result in the formation of a task-set that can be used to guide performance until it is no longer needed, and the task can be completed from memory alone. A task schema can therefore be considered a set of broad strategic parameters that can be applied to multiple tasks with a common structure (e.g., stimulus → classification → response), whereas a task-set is a specific group of

parameters that are applied to a relatively limited context (e.g., 3, 5, 7 → odd number → left index finger key press). Automatic execution is thought to be context-specific (e.g., Logan, 1988) and therefore relies on specific instances (e.g., 7 → left index finger key press). Although there is a wealth of research on the controlled execution and automatic execution stages of learning, there is little research that has focused specifically on the formation stage and the transfer of task schemas between tasks with a common structure.

Although we introduce the term 'task schema' here, the notion that learning might not be limited to associations between representations but can also include the underlying structure of a set of related tasks is not novel (e.g., Ableson, 1981; Maslow, 1949; Thorndyke & Yekovich, 1980) and has even been the subject of more recent research (e.g., Cooper & Shallice 2006; Braun et al., 2010). Although there are a few contemporary examples of research that specifically investigate this kind of learning (see below), remarkably little is known about the details of how task schemas are represented, their architecture, or how they orchestrate learning a novel task. The central aim of the present study was to contribute to this literature by conducting a relatively large-scale, systematic investigation of these issues with an eye to develop our understanding of an important aspect of human learning and transfer that remains poorly understood.

Learning and transfer

Research on learning and automaticity has a long tradition of using tasks that rely on the formation of simple stimulus-response (S-R) associations (e.g., A → left hand key press, B → right hand key press). However, there is a recent recognition of the limits inherent in such a narrow view (e.g., Hazletine & Schumacher, 2016; Schumacher & Hazletine, 2016). A growing body of research has been investigating learned associations between other kinds of representations such as categories of stimuli and/or responses (for a relevant review, see

Henson et al., 2014). For example, Horner and Henson (2009, 2011) asked participants to classify pictures of everyday items in a study-test design. By manipulating repetitions of the stimulus (picture, word), the classification (e.g., larger than a shoebox, larger than a wheelie bin), the decision (yes, no) and the action (left/right key press) between study and test phases, they were able to demonstrate that stimuli and responses can be represented at different levels of abstraction. For example, a stimulus can be represented as a specific exemplar or a semantic/perceptual category of stimuli whereas a response can be represented motorically or semantically (for some other examples of research consistent with this idea see, e.g., Dennis & Perfect, 2013; Longman et al., 2018, 2019, 2020; Moutsopoulou & Waszak, 2012, 2013; Pashler & Baylis, 1991; Pfeuffer, et al., 2017; Wills et al., 2006). The common conclusion from this body of research is that existing associations between the learned material and other instances from the same stimulus/response category can expedite learning when those instances are introduced during the controlled execution stage. For example, learning that the letters A, E, and I require a left-hand response might result in speeded responses to untrained instances from the same category (i.e., the vowels O and U) due to existing associations between these instances as part of a broader category structure.

However, learning need not be limited to the associations formed between (sets of) stimuli, categories, and/or responses. Learning can also be observed when practiced task components are combined in novel ways. For example, Cole and colleagues (2011; see also Cole, et al., 2010) developed a paradigm that allowed them to give participants training on a set of 12 task components (rules) that could be combined to make 64 unique tasks. On each run of trials, a unique task (a unique combination of the task components) was cued by instructing the participant in the semantic classification (e.g., ‘sweet’), the comparison (e.g., ‘same’), and the response (e.g., ‘left index finger’) which should be applied to the subsequent word pairs (e.g., ‘grape’ and ‘apple’). Four of the possible tasks (unique combinations of the

12 components) were extensively practiced prior to the experimental session. Cole et al. found that initial performance was better for the tasks introduced for the first time in the experimental session (which were made up of practiced components combined in unique ways) relative to the tasks introduced for the first time during the training session (which were made up of unpractised components). That is, participants were able to transfer their prior experience with individual task components (e.g., the semantic classification "sweet") to novel tasks that used the same components in unique combinations thereby expediting the early stages of learning (for a similar theory based on computational modelling, see Taatgen, 2013).

Finally, flexible behaviour need not necessarily rely on combining pre-learned task components in novel ways. People can also learn about the task structure itself. For example, learning about the task has received some attention in the (hierarchical) reinforcement learning literature (as reviewed by Bhandari & Badre, 2018). In these studies, participants create a task schema, or internal task model, on the fly which can be adapted to fit novel tasks with a similar structure (for similar theories see, e.g., Bhandari & Duncan, 2014; Braun et al., 2010; Collins & Frank, 2013; Schmidt et al., 2015; Verbruggen et al., 2014). For example, Bhandari and Badre found both positive and negative transfer of working memory (WM) gating policies (i.e., task strategies) across unique tasks where optimum performance relied on either efficiently updating the contents of WM or biasing attention toward particular items stored therein. That is, they found evidence that task strategies readily transferred to novel tasks that used unique stimuli, even if the strategies were not appropriate to the current context. More recently, Pereg et al. (2021) observed a systematic improvement in performance through a session consisting of multiple unique simple two-choice S-R learning tasks. Such an improvement in performance cannot easily be explained by transfer of learned associations, or compositionality, but can easily be explained by participants learning about

the structure of the tasks themselves and applying the relevant (learning) strategy to other tasks with a common structure.

To summarise, learning can include structural information about groups of tasks where performance can be optimized by using a common (learning) strategy (e.g., Bhandari & Badre, 2018). Such information can be readily applied to unique tasks that share a common structure thereby expediting learning (e.g., Pereg et al., 2021). Learning novel tasks can also be expedited by recombining the elements of practiced task-sets into unique combinations (e.g., Cole et al., 2011). However, it is still not known whether structural representations can also be broken down into their constituent components, and whether these components can be recombined in novel ways to expedite learning novel tasks that share *some* common structural elements with the training tasks (e.g., the same *kinds of* stimuli, classifications and/or responses), but none of the surface features (e.g., the *specific* stimuli, classifications and/or responses). That is, the core aim of the present study was to determine whether *strategic* elements of practiced tasks with a common structure could be combined in novel ways to expedite the early stages of learning a unique task via the rapid formation of a novel task schema based on components of an existing schema. And if so, which specific strategic elements (or combinations of elements) are most likely to transfer between tasks to expedite learning.

Experiment 1

Experiment 1 was an exploratory study designed as a first attempt to systematically investigate the architecture of a task schema by dividing a class of dot-pattern classification tasks (first introduced by Longman et al., 2018) into perceptual, categorization, and motoric components and providing groups of participants with training on all/some/none of those

components before comparing their performance in equivalent dot-pattern classification tasks at test.

In the *full schema* training condition, participants performed three distinct dot-pattern classification tasks (each task used completely novel stimuli, categories and responses) before completing a fourth dot-pattern classification task in the final test block. In a *control* training condition, participants had to indicate whether some simple mathematical formulae were true or false by making a left/right key press response (i.e., none of the test task elements could be practiced during the training phase). Two other groups were able to gain experience with the *motoric*, and *perceptual + motoric* components of the test task. A final group (*categorization + motoric)* were given the opportunity to perform a different set of classification tasks but use the same kinds of motor responses as the test task. By comparing performance from the start (and end) of the test block (using RT and accuracy measures from the first/last 3 repetitions of each stimulus in a block) across the training groups we could determine which task elements directly contributed to subsequent performance/learning in a novel task, and whether any benefits observed in the early stages of learning were maintained after a short spell of practice.

Note that the *control* training tasks were designed to control for extraneous variables such as fatigue, and time engaged in a broadly similar experimental task prior to the test block. However, it could be argued that participants in this group might have benefited from performing any task for a comparable duration to the participants from the other groups prior to the test phase (e.g., they might have become accustomed to being in the laboratory, concentrating on a simple task, or engaging with the same experimental context). For this reason, we also compared performance from Block 1 for the *full schema* training group (who performed a unique dot-pattern classification task without any prior experience of the experimental context) to test performance from each training group in each experiment.

These *alternative control* performance data are plotted in the relevant figures, but their analysis is reported in the Supplementary Materials. The findings from the analyses suggest that the *control* training regime was, if anything, a conservative estimate of baseline performance given that performance was significantly worse in the *alternative control* relative to the *control* group in all experiments.

We predicted that the *full schema* training group would outperform the *control* training group in the early stages of the test task. Such a finding would provide direct evidence that experience with tasks that share a common structure with the test task can expedite learning via the application of an existing task schema (cf. Bhandari & Badre, 2018; Pereg et al., 2021).

The purpose of a task schema (learning strategy) is to aid in the formation of a useful task-set that can be applied to the current context. That is, task schemas are only active in Chein and Schneider's (2012) formation stage of learning that has a duration in the order of seconds. We therefore predicted that any performance benefits associated with task schema transfer would be relatively short lived (i.e., only detectable within the first few trials/stimulus repetitions of learning a novel task – the formation stage). If any performance differences remained reliable at the end of the test block (during the controlled execution stage), then the benefits might not be explained by transfer of task schemas alone.

Critically, we also predicted that the *motoric*, *perceptual + motoric*, and/or the *categorization + motoric* training groups would outperform the *control* training group in the early stages of the test block. Such a finding would indicate that a task schema can be divided into distinct components, each of which might expedite learning in a task that shares *some* structural elements with the training tasks (i.e., something akin to compositionality of structural task components; cf. Cole et al. 2011).

Method

Participants

The final sample included a total of 200 individuals (40 per training group) who were recruited via the University of Exeter participant pool. Demographic data for each group can be found in Table 1. Each participant was paid £4 or was awarded partial course credit for participating in a session that lasted around 30 minutes. N was set prior to testing (and was selected based on the design of Longman et al., 2018). A power analysis using G*Power (Faul et al., 2007) determined that, when N = 40 in each group, it was possible to detect medium-large sized differences (Cohen's d = 0.78) with 80% power with an Alpha of 0.005 (0.05 Bonferroni corrected for 10 one-tailed comparisons). This experiment was approved by the local research ethics committee at the School of Psychology, University of Exeter. Written informed consent was obtained after the nature and possible consequences of the study were explained.

Table 1
Demographic Data for Each Training Group in Experiment 1.

Training group	Mean age (SD)	Gender	Dominant hand
Full schema	20.2 (3.0)	Male: 8 (20.0%) Female: 32 (80.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 35 (87.5%) Left: 5 (12.5%)
Motoric	19.7 (3.9)	Male: 6 (15%) Female: 34 (85.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 38 (95.0%) Left: 2 (5.0%)
Perceptual + Motoric	22.2 (6.5)	Male: 10 (25.0%) Female: 30 (75.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 36 (90.0%) Left: 4 (10.0%)
Categorization + Motoric	20.4 (6.4)	Male: 6 (15.0%) Female: 34 (85.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 37 (92.5%) Left: 3 (7.5%)
Control	19.6 (3.8)	Male: 9 (22.5%) Female: 31 (77.5%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 34 (90.0%) Left: 6 (10.0%)

Apparatus, Stimuli, and Responses

Stimuli were presented on a 21.5-inch iMac using Psychtoolbox (Brainard, 1997). As shown in Figure 1 the stimuli used in the dot-pattern classification tasks of the *full schema* training condition and the test phases consisted of patterns of five black dots (diameter = 0.5 cm) presented at a pseudo-random location in a larger array (18 x 18) of small grey dots (diameter = 0.25cm, distance between adjacent dots = 0.75 cm), which was itself surrounded by a black square (side = 15 cm, thickness = 2 ptu). Three of the black dots were positioned at locations within the category template and the remaining two were positioned at locations directly adjacent to the template. Thus, category membership of a given stimulus (eight per block, four per category) was determined by its overall similarity to the category templates (two per block, pseudo-randomly selected from a set of five pairs).

The stimuli for the *motoric* training were the relevant response code (two per block, order randomized) presented centrally in black Arial font (size 30). The *perceptual + motoric* training tasks used the same stimuli as the dot-pattern classification tasks, but they were presented either to the far left or far right of the grid (pseudorandomized, with each stimulus appearing on either side of the grid equiprobably; vertical position was randomized).

The stimuli for the *categorization + motoric* training tasks were drawings of space ships, aliens or flowers (one category per block, order randomized) that could differ on four binary dimensions (Space ship: shape of ship, shape of window, shape of wings, number of rockets; Alien: shape of antennae, number of eyes, number of fingers, shape/number of teeth; Flower: shape of head, diameter of centre, shape of leaves, shape of pot). All individual stimuli from these picture classification tasks are deposited on the Open Science Framework data repository: https://osf.io/t9whg/. The two category templates used in each experimental block differed on all four dimensions and therefore had no critical features in common with each other. Each stimulus (eight in total, four per category) differed from the relevant

template on one dimension (a different dimension for each stimulus, no single dimension was 100% predictive of category membership; for a similar category structure, see Medin, Wattenmaker, & Hampson, 1987) and was presented surrounded by a black square identical to the square surrounding the grid in the dot-pattern classification blocks. Like in the dot-pattern classification tasks, category membership of each stimulus in the picture-classification tasks was determined by overall similarity to the templates.

The stimuli for the *control* training condition were simple mathematical formulae (eight per block, four true, four false) presented centrally in a randomized order, in black Arial font (size 30).

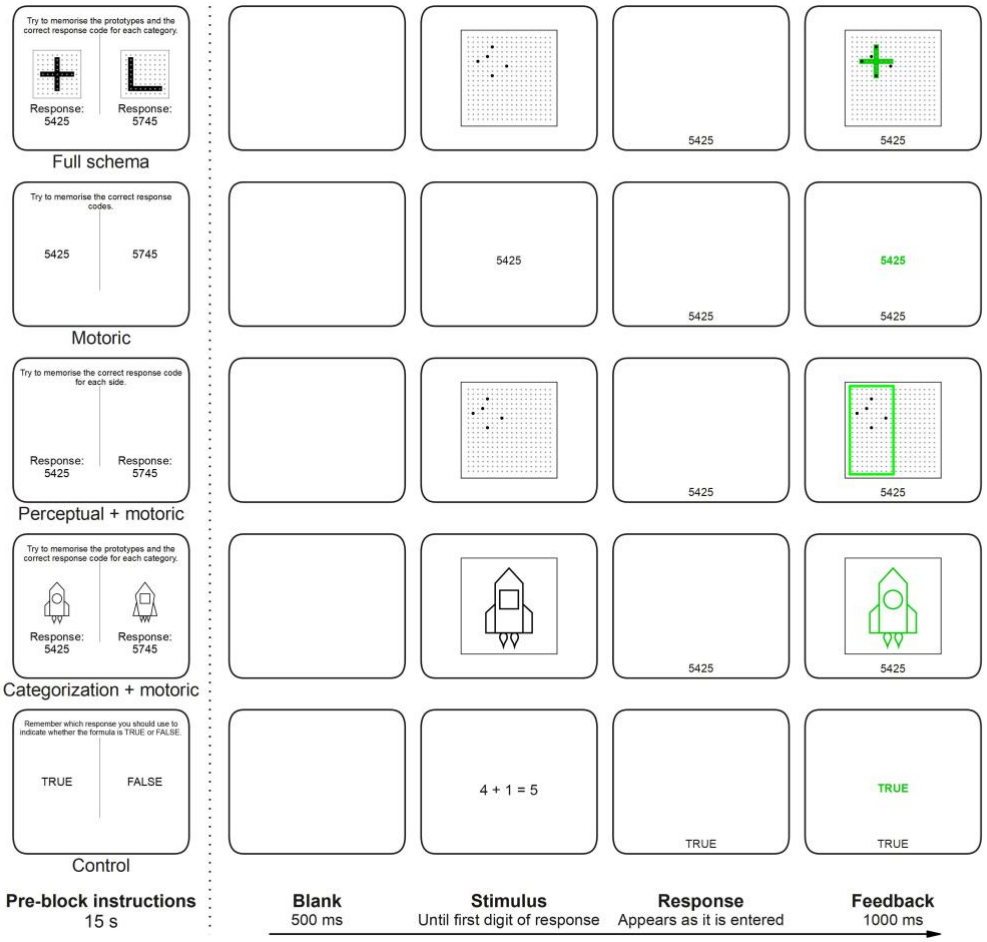
On each trial in all conditions except for the control training blocks, the participant responded by entering a four-digit response code (two codes per block) with their right index finger (participants were asked to use their preferred hand, all chose to respond with their right hand) using the numeric keypad part of a standard keyboard. The response codes all started and ended with '5' and the intervening digits were always on adjacent keys to equate the difficulty of entering each code (e.g., 5235, 5425). The codes used to respond in each block were selected pseudo-randomly (from a set of eight pairs) to equate the difficulty of each task. In the control training blocks, the participant responded by pressing either the 'F' or 'J' keys on a standard keyboard with their left and right index fingers respectively.

Procedure

The procedure for a single trial in each condition is presented in Figure 1. Each trial started with a blank screen displayed for 500 ms, after which the stimulus appeared. In all tasks that used four-digit responses, the stimulus remained visible until the participant entered the first digit in the response code. The response code then appeared at the bottom of the screen as it was typed in (black Arial font size 30). In the control training condition, which

used single left/right key press responses, the stimulus remained visible until any response was made, and the given response ('TRUE' or 'FALSE') was displayed at the bottom of the screen for 200 ms (black Arial font size 30). In all conditions, immediate feedback was provided on every trial and was visible for 1000 ms (see Figure 1).

Figure 1
Pre-block instructions screen (left) and timeline of a single trial (right) for each training condition in Experiment 1.



Note: The relevant part of the feedback screen was presented in either green (correct) or red (error). The pre-block instructions encourage the participant to memorize the relevant task components while they are visible. In the control training blocks, the given response was presented for 200 ms.

The experiment consisted of four experimental blocks of 128 trials each divided into two mini-blocks of 64 trials, with a short break between mini-blocks. At the start of the

session, participants were informed that the first three experimental blocks were an extended training phase and that the fourth and final experimental block was the test phase in which they would have to perform a dot-pattern classification task. All groups completed three unique tasks in the training phase (one task per block). For example, the *full schema* training group completed three dot-pattern classification tasks with unique stimuli, categories, and responses where the task was to indicate which of the two categories the displayed stimulus belonged by entering the relevant response code, whereas the *motoric* training group simply had to enter three unique pairs of response codes (see Figure 1). The test task was equivalent for all groups. Participants had to indicate which of two perceptual categories the displayed dot-pattern belonged to by entering the relevant response code. The stimuli, categories, and responses used in the test task were always unique. In all conditions, participants first received identical instructions regarding the dot-pattern classification task before receiving (condition-specific) instructions regarding the training phase. A reminder about the procedure of the dot-pattern classification task (printed in text, visible for 30 seconds) immediately preceded the pre-block instructions for the test phase. Prior to the start of each mini-block, an instruction screen displayed all relevant information necessary to perform the current task (see Figure 1). Each mini-block ended with a feedback screen displaying the mean RT and proportion of errors for that mini-block. The pre mini-block instructions and the feedback screen were both visible for 15 seconds.

Dependent variables and Analyses

All data processing and analyses were performed using R (R Core Team, 2022). Bayes Factors were calculated using the BayesFactor package (Morey et al., 2022) and plots were drawn using ggplot2 (Wickham, 2022). All raw data files, R scripts, and the experiment

scripts used for data collection from all experiments are deposited on the Open Science Framework data repository (https://osf.io/t9whg/).

In each training condition, trials with RT (first digit) <100 ms (*full schema* = 0.08%, *motoric* = 0.08%, *perceptual + motoric* = 0.02%, *categorization + motoric* = 0.07%, *control* = 0.02%) and trials with RT (last digit) >5000 ms (*full schema* = 0.98%, *motoric* = 0.48%, *perceptual + motoric* = 0.49%, *categorization + motoric* = 2.35%, *control* = 0.57%) were omitted from all analyses. Error trials were also omitted from RT analyses. Any participants who had <50% of the maximum possible observations in at least one mini-block following the above data cleaning procedures were replaced (*full schema* = 5; *motoric* = 3; *perceptual + motoric* = 2; *categorization + motoric* = 9; *control* = 2). The latency of the final digit in the response code (i.e., the time at which the full response had been entered) was used for all response latency analyses[1] (Longman et al., 2018). Performance from all stimuli in each block were pooled.

Two critical measures were used to compare test[2] performance between the five (training) groups. As noted above, we were especially interested in performance during the earliest stages of learning a new task. Our first measure, *initial performance*, therefore took the average performance during the first three repetitions of each stimulus in the test block[3]. Our second measure, *final performance*, took the average performance during the final three repetitions of each stimulus in the test block. This measure was used to determine the extent

[1] Note that the response in the training blocks for the control group was a single key press (though it was a complex four-digit response, as in the other conditions, at test). The latency of the single response was used for the data cleaning procedures. The same boundaries were used as for the other conditions.

[2] Analysis of the training performance data was not directly relevant to the research question so it can be found in the Supplementary Materials.

[3] Note that each bin included the data from a maximum of 24 trials (3 repetitions of 8 stimuli).

to which any differences between groups observed in initial performance were still observable at the end of the test block.

Because this was an exploratory experiment, the performance from each training group was compared to all other training groups. This strategy allowed us to systematically investigate the benefits of each training condition relative to the others. Thus, for each measure and dependent variable (RT and error rate), we performed a separate set of independent-samples *t*-tests. We applied separate Holm-Bonferroni corrections[4] for each set of comparisons. We also performed equivalent Bayesian analyses (using the default JZS prior of .707) and calculated effect sizes (Hedges's *g*; Lakens, 2013). For all Bayesian analyses only the BF_{10} is reported (i.e., the Bayes Factor for evidence in favour of the alternative hypothesis). Bayes factors are interpreted according to the protocol outlined by Schönbrodt and Wagenmakers (2017): BF > 100 = extreme evidence for H_1, BF 30-100 = very strong evidence for H_1, BF 10-30 = strong evidence for H_1, BF 3-10 = moderate evidence for H_1, BF 1-3 = anecdotal evidence for H_1, BF 1 = no evidence, BF 0.333-1 = anecdotal evidence for H_0, BF 0.01-0.333 = moderate evidence for H_0, BF 0.033-0.1 = strong evidence for H_0, BF 0.001-0.033 = very strong evidence for H_0, BF < 0.001 = extreme evidence for H_0.

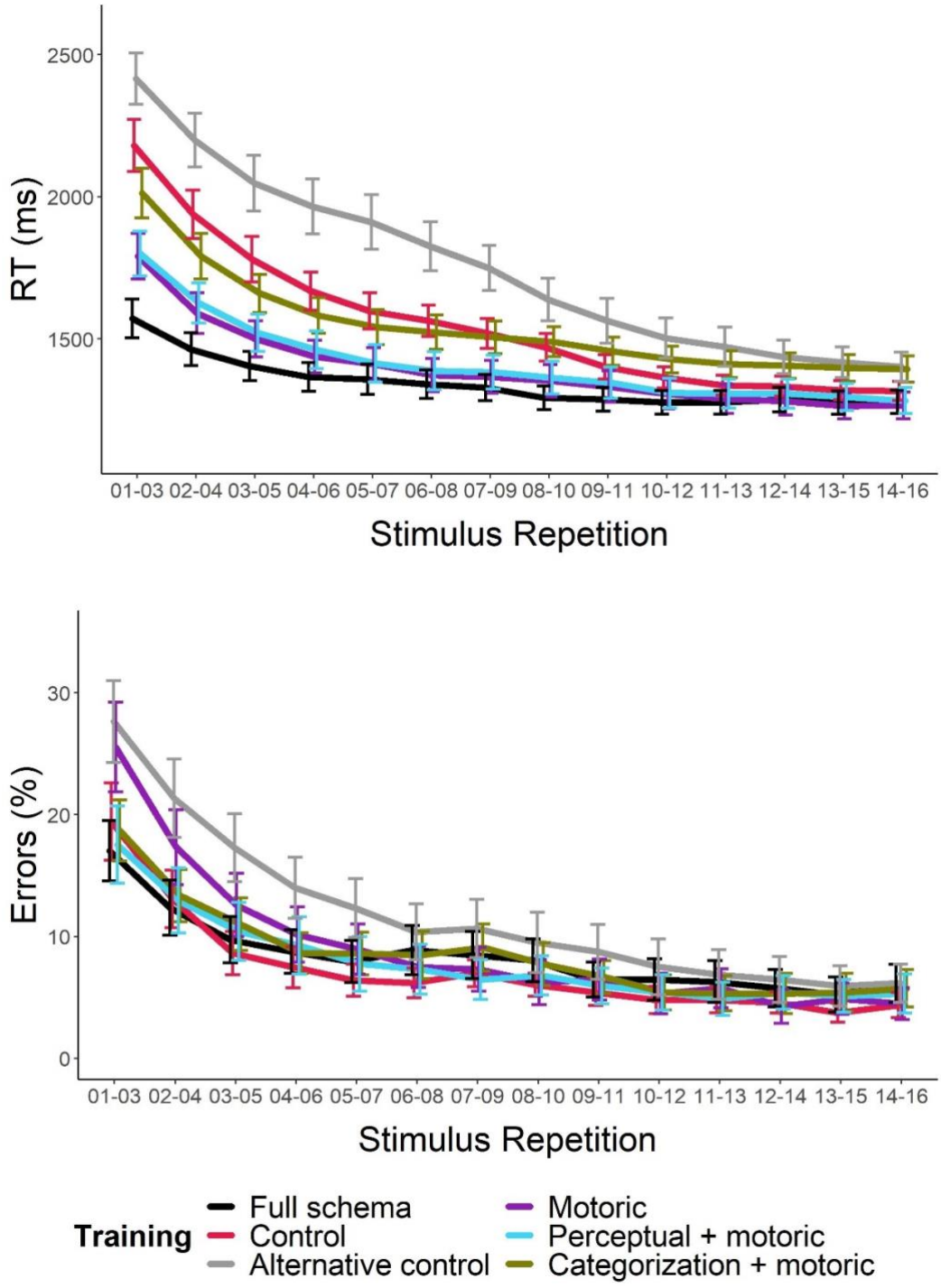
Results and Discussion

Mean RTs and proportion of errors from the test block are plotted as a function of training group and stimulus repetition in Figure 2. Note that rolling means were used to visualise the data, but all analyses used only the data from the first and last points on the curves. For brevity, the results from all planned contrasts comparing test performance across

[4] Holm-Bonferroni correction controls for family-wise error rate by adjusting the alpha value for each individual comparison.

the training groups are reported in the Supplementary Materials and only the main findings of interest are summarized here.

Figure 2
Mean RTs (left) and proportion of errors (right) from the test blocks of Experiment 1 plotted as a function of training condition and stimulus repetition.



Note: Error bars show between-subjects difference adjusted 95% confidence intervals.

Initial test RTs were significantly faster in the *full schema* group relative to all the other groups (ps ≤ 0.012, BFs > 3.8) indicating that training with multiple dot-pattern classification tasks resulted in improved initial test performance relative to the other training conditions. Note that the *t*-test comparing initial test performance in the *full schema* and *motoric* training groups did not survive correction for multiple comparisons, but the Bayesian analysis (for which the Holm-Bonferroni procedure does not apply) found moderate evidence that initial RTs were faster in the *full schema* training group than in the *motoric* training group.

Initial RTs were significantly slower in the *control* training group relative to all other groups (ps < 0.001, BFs > 120) with the exception of the *categorization + motoric* group (p=0.105, BF=0.740) indicating that training with some/all task components improved test performance relative to a condition where the training was (largely) irrelevant to the test task. None of the remaining comparisons of initial RTs survived correction for multiple comparisons suggesting that practice with any subset of strategic task components relevant to the test task was broadly equivalent. That is, training with any of the strategic task components relevant to the test task improved initial RTs relative to the control but training with the full schema provided an additional benefit.

Note that all training groups gained experience with the motoric component of the test task. Importantly, the *motoric* training group, who only gained experience with this component, had similar initial test performance to the other groups who gained experience with a subset of the strategic task components relevant to the test task. This finding suggests that training with the motoric component of the test task was an important contributor to initial test performance in the present paradigm. However, training with the full schema still gave participants some advantage in initial test performance (note that the Bayesian analysis

found moderate evidence that the *full schema* group outperformed the *motoric* group) suggesting that experience with other components also contributed to the improvement in initial test RTs for this training group.

None of the planned comparisons performed on the initial proportion of errors from the test block survived correction for multiple comparisons (ps ≥ 0.020) indicating that response accuracy was largely unaffected by the differences in training between groups. Finally, none of the planned comparisons on the final performance measures survived correction for multiple comparisons for either RTs (ps ≥ 0.019) or errors (ps ≥ 0.246) indicating that any differences in the initial performance across the groups had largely been resolved by the end of the test blocks.

In summary, the results from Experiment 1 are consistent with the theory that transfer of task schemas can expedite the early stages of learning a new task that shares a common structure with the training tasks. Importantly, the present findings also support the notion that a task schema can be divided into distinct strategic task components, some of which can offer benefits to subsequent learning in a task that shares a subset of the structural task components practiced in the training tasks. Although training with the full schema was associated with the largest benefits to subsequent performance, training with the motoric component was also an important contributor to initial RTs in the test task. However, the benefits of motoric training might be especially prevalent in the present design where the required response was a 4-digit code (we return to this issue in the General Discussion). Nonetheless, any performance benefits observed during the early stages of the test block were largely eliminated by the end of the block suggesting that the benefits were rather short lived. The latter conclusion is consistent with the notion that the early benefits to performance observed in the present experiment were due to task schema transfer which expedited the formation of a stable task-

set. Once the relevant task-set had been formed/implemented, the performance benefits of task-schema transfer were largely eliminated.

Experiment 2

Experiment 1 was an initial exploratory investigation into the architecture of task schemas. Although the outcomes were informative to a degree, there remain several unanswered questions and possible optimizations of the design. The primary aim of Experiment 2 was to address these issues.

For example, some training conditions were not included in the design of Experiment 1. Although one group were able to gain experience of the motoric element of the test task (independently of the remaining structural components), no groups were given training with the perceptual or categorization elements independently of the remaining elements of the test task. An additional group in Experiment 1 gained experience with the perceptual and motoric elements of the test task, but no groups were trained with the perceptual and categorization elements, or with the same kinds of categorization task and responses used in the test task. Finally, the *categorization + motoric* training in Experiment 1 used a different kind of categorization task that did not allow participants to gain experience with classifying stimuli by comparison to a geometric perceptual category template like those used in the test task. Experiment 2 was designed primarily to address these omissions in the training groups included in Experiment 1. To this end, Experiment 2 included a total of eight training groups – *perceptual*, *categorization*, *motoric*, *perceptual + categorization*, *perceptual + motoric*, *categorization + motoric*, *full schema*, and *control* (see below for full details on each of these). The inclusion of the additional training groups allowed a fuller analysis of the benefits of training with single structural task components and all combinations of two-component training regimes relative to the full schema and control conditions (see Analyses section). The

design also allowed for a full factorial analysis of the performance data from the test block as a function of whether the training included perceptual, categorization, and/or motoric components and a direct analysis of test performance as a function of the number of relevant components included in the training tasks.

The rationale for the duration of the training phase in Experiment 1 (three 128-trial blocks) was based on the observation that most of the performance benefits to be gained by transfer of task schemas observed by Longman et al. (2018) were found across the first three training blocks. However, theories of learning and automaticity (e.g., Chein & Schneider, 2012; Logan, 1988) suggest that, as new skills become more automatic, performance relies less on algorithmic processes (rules) and more on memory-based processes (associations) – i.e., that expertise is context-specific, or that novice performance is more likely to transfer between tasks. Thus, one might suppose that a longer training phase with a broader range of training tasks, but a shorter training block for each task might enhance the effects of transfer. For this reason, the training phase in Experiment 2 was extended to include five 48-trial[5] blocks, each of which was associated with a unique task. It was predicted that this manipulation to the design would enhance the differences in test performance between groups thereby allowing for more subtle differences to be measured.

Experiment 2 was run online because of restrictions associated with the Covid-19 pandemic. However, a benefit of doing so was that the population was not limited to the

[5] Note that inspection of Figure 1 suggests that most of the initial test RT differences between groups in the present Experiment 1 had been resolved by the 8th stimulus repetition indicating that a 48-trial test block should be enough to observe the differences of interest.

student population from a particular university and the sample was instead drawn from a much more diverse pool thereby improving the generalizability of the results.

Method

Participants

The final sample included a total of 640 individuals (80 per training group) who were recruited via 'Prolific' (an online participant pool) and paid £3.15 for a 25-minute session. The only criteria used to filter participants at this stage was that they identified as being fluent in English and were using a desktop or laptop computer with a numeric keypad on the right-hand side of the keyboard. Demographic data for each group can be found in Table 2. N was set prior to testing and was selected based on a compromise between increasing power relative to Experiment 1, and access to fair payment for participants. G*Power determined that, when N = 80 in each group, it was possible to detect medium-large sized differences (Cohen's d = 0.56) with 80% power with an Alpha of 0.004 (0.05 Bonferroni corrected for the 13 one-tailed comparisons we chose *a priori*, see below). This experiment was approved by the local research ethics committee at the University of the West of Scotland. Participants indicated their informed consent by ticking a box on screen after the nature and possible consequences of the study had been explained. All participants were debriefed on-screen at the end of the session.

Table 2

Demographic Data for Each Training Group in Experiment 2.

Training group	Mean age (SD)	Gender	Dominant hand
Full schema	28.1 (11.0)	Male: 52 (65.0%) Female: 28 (35.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 74 (92.5%) Left: 6 (7.5%)
Perceptual	27.5 (9.7)	Male: 59 (73.8%) Female: 20 (25.0%) Diverse: 0 (0.0%) Prefer not to say: 1 (1.3%)	Right: 74 (92.5%) Left: 6 (7.5%)
Categorization	25.4 (7.1)	Male: 50 (62.5%) Female: 28 (35.0%) Diverse: 2 (2.5%) Prefer not to say: 0 (0.0%)	Right: 70 (87.5%) Left: 10 (12.5%)
Motoric	25.9 (8.2)	Male: 48 (60%) Female: 31 (38.8%) Diverse: 1 (1.3%) Prefer not to say: 0 (0.0%)	Right: 71 (88.8%) Left: 9 (11.3%)
Perceptual + Categorization	24.5 (5.6)	Male: 49 (61.3%) Female: 31 (38.8%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 70 (87.5%) Left: 10 (12.5%)
Perceptual + Motoric	26.1 (8.0)	Male: 50 (62.5%) Female: 28 (35.0%) Diverse: 1 (1.3%) Prefer not to say: 1 (1.3%)	Right: 72 (90.0%) Left: 8 (10.0%)
Categorization + Motoric	26.0 (7.6)	Male: 56 (70.0%) Female: 24 (30.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 73 (91.3%) Left: 7 (8.8%)
Control	27.0 (10.2)	Male: 48 (60.0%) Female: 32 (30.0%) Diverse: 0 (0.0%) Prefer not to say: 0 (0.0%)	Right: 69 (86.3%) Left: 11 (13.8%)

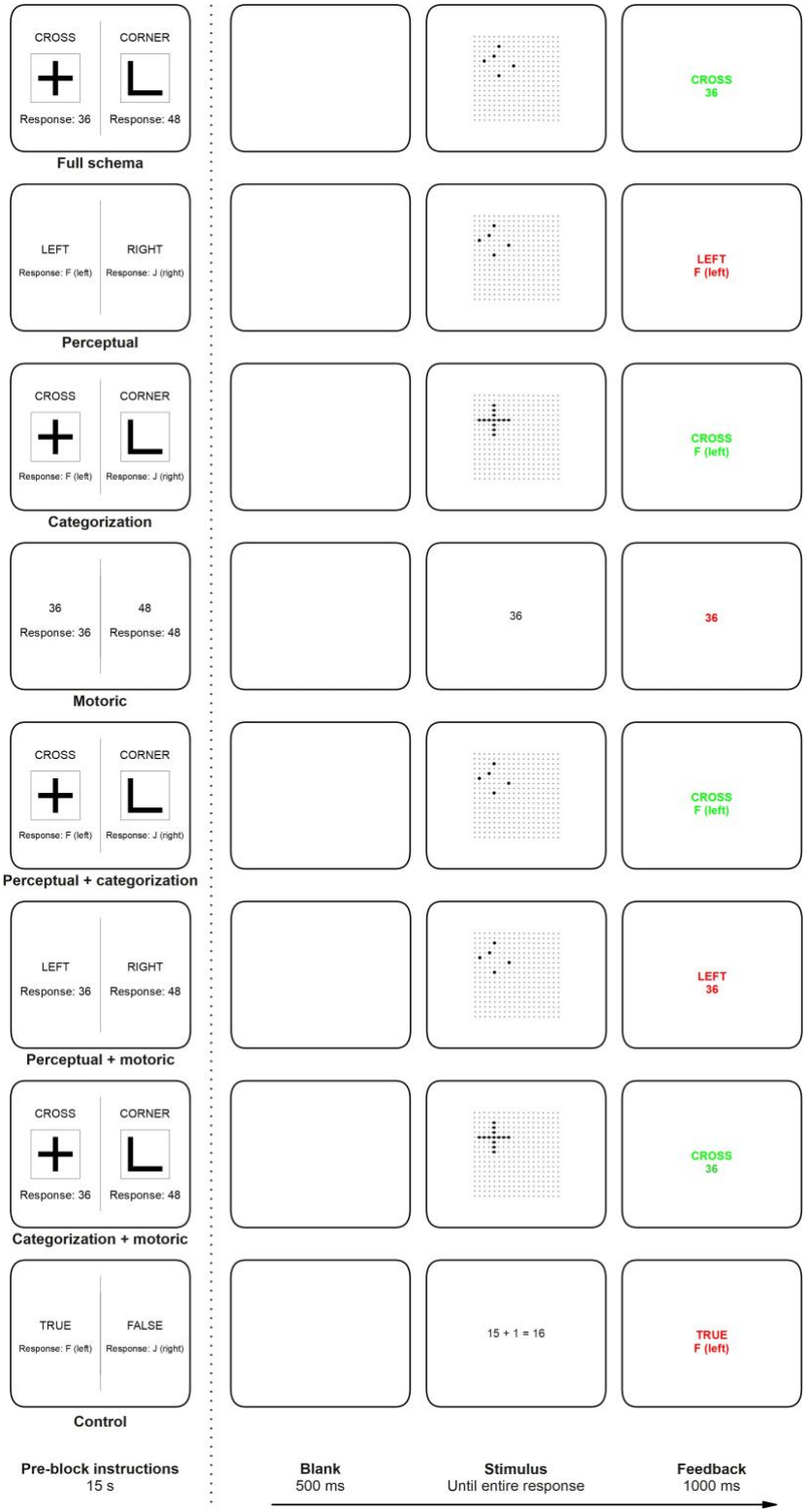
Materials, Design, and Procedure

Experiment 2 was run online, so it was not possible to determine the precise hardware used by each participant. However, individuals could only participate by using a laptop or desktop computer with a numeric keypad on the right-hand side of the keyboard (not a smartphone or a tablet) connected to the internet. All tasks were coded using JavaScript (jsPsych; De Leeuw, 2015) and were hosted by Pavlovia.org. The stimuli, responses, and general structure of the session were based on those used in Experiment 1, but there were several important differences. The session consisted of six blocks of 48 trials each (4 stimuli from each category were presented 6 times each per block). Participants were informed that

the first five blocks were the training phase, and the final block was the test phase. Participants from all conditions first received identical instructions regarding the test task before receiving (condition-specific) instructions regarding the training phase. A reminder about the procedure of the dot-pattern classification task (printed in text, visible for 60 seconds) immediately preceded the pre-block instructions for the test phase. A unique task was performed in each block throughout the entire session. The nature of the tasks performed during the training phase depended on the group, but all participants performed a unique dot-pattern classification task at test similar to those performed in the full schema training blocks (see Figure 3).

Figure 3

Pre-block instructions screen (left) and timeline of a single trial (right) for each training condition in Experiment 2.



Note: The relevant part of the feedback screen was presented in either green (correct) or red (error). The pre-block instructions encourage the participant to memorize the relevant task components while they are visible.

As shown in Figure 3, the stimuli used in the dot-pattern classification tasks of the *full schema* and the *perceptual + categorization* training tasks as well as the test phase for each group were identical to those used in the test phase of Experiment 1. As in Experiment 1, the *perceptual + motoric* training tasks used the same stimuli as the dot-pattern classification tasks, but they were presented to the far left or far right of the grid (pseudorandomized, with each stimulus appearing on either side of the grid equiprobably; vertical position was randomized). The same stimuli were also used in the present *perceptual* training tasks. The stimuli for the *categorization* and *categorization + motoric* training tasks were the relevant category templates drawn at a random location in the grid using arrows pointing up, down, left or right (equivalent size to the black dots used in the dot-pattern classification tasks). In these conditions, the participants could therefore learn to identify the relevant category by comparing the stimulus to an internal representation of the category template but could not gain experience of comparing the templates to the dot-pattern stimuli used in the test task. That is, they could learn about the kinds of categories, but not the kinds of stimuli used in the test task. The use of arrows instead of dots for these stimuli meant that each category could include four unique exemplars, but the differences between them would be minimal thereby encouraging participants to simply identify the category template, rather than having to identify the category to which the distorted stimulus belonged. As in Experiment 1, the stimuli for the *motoric* training tasks were the relevant response code (two per block), and the stimuli for the *control* training tasks were simple mathematical formulae (eight per block, four true, four false). Both of the latter were presented centrally in a randomized order in black Arial font.

On each trial in the *motoric*, *perceptual + motoric*, *categorization + motoric*, and *full schema* training blocks as well as in the test tasks for all groups, participants were asked to

respond by entering a two-digit response code equivalent to the four-digit response codes used in Experiment 1 with the first and last digits (5) omitted from the code. Participants were asked to respond using only one finger from their right hand (the instructions suggested the index finger but emphasized the importance of using only one finger to make each response) using the numeric keypad part of a standard keyboard. Although none of the codes included the number 5, participants were encouraged to start each trial with their finger in this central position on the numeric keypad. The codes used to respond in each block were selected pseudo-randomly (from a set of 8 pairs) to equate the difficulty of each task. Note that the coding of the experiment meant that it was not possible for participants to respond using the number keys along the top of the keyboard. This precaution was introduced to ensure that all participants used the numeric keypad to enter the response codes. In the remaining training blocks, the participant responded by pressing either the 'F' or 'J' keys on a standard keyboard with their left and right index fingers respectively.

Otherwise, the procedure for Experiment 2 was identical to that for Experiment 1 with two additional exceptions. First, the response code did not appear on the screen as it was typed in by the participants. Instead, the stimulus remained visible until the entire response code was entered. As in Experiment 1, immediate feedback was then displayed for 1000 ms. Second, on trials where the stimulus was presented within a grid of small grey dots, this grid was not surrounded by a black square.

Analyses

In each training condition, trials with RT (first digit) <100 ms (*full schema* <0.01%, *perceptual* = 0.02%, *categorization* <0.01%, *motoric* = 0.02%, *perceptual + categorization* <0.01%, *perceptual + motoric* = 0.01%, *categorization + motoric* = 0.00%, *control* <0.01%) and trials with RT (last digit) >3000 ms (*full schema* = 2.05%, *perceptual* = 0.63%,

categorization = 0.72%, *motoric* = 0.55%, *perceptual + categorization* = 0.79%, *perceptual + motoric* = 1.00%, *categorization + motoric* = 1.25%, *control* = 1.33%) were omitted from all analyses. Error trials were also omitted from RT analyses. Any participants who had <50% of the maximum possible observations in at least one block following the above data cleaning procedures were replaced (*full schema* = 9, *perceptual* = 8, *categorization* = 2, *motoric* = 3, *perceptual + categorization* = 5, *perceptual + motoric* = 5, *categorization + motoric* = 5, *control* = 10). (Note that participants were informed that they would not be eligible for payment if they achieved less than 75% accuracy in any single block, and that responses slower than 3000 ms would be counted as errors in this calculation).

The same two critical measures used in Experiment 1 were used again in Experiment 2. However, because Experiment 2 had a full factorial design, it was possible to conduct several analyses in addition to the planned contrasts comparing test performance between training conditions. First, to determine whether training with the perceptual, categorization, and/or motoric components of the test task independently contributed to test performance, we ran a separate Perceptual (component was present/absent from training) by Categorization (component was present/absent from training) by Motoric (component was present/absent from training) between-subjects ANOVA on each dependent variable for each measure. Bayes factors and effect sizes (generalized eta squared) were also calculated for all relevant effects/interactions. Bayes factors were calculated with the BayesFactor package, using the default JZS prior (.707; Morey, Rouder, & Jamil, 2015). To reduce the number of model comparisons, a subtraction approach was used (see Morey et al., 2015). That is, the full model including all effects and interactions was compared to the full model minus the effect or interaction of interest. When this approach is used, Bayes factors < 1 indicate that removing the effect/interaction reduces the fit of the model (i.e., it is a contributor to the fit of the full model). Conversely, when the Bayes factor remains > 1, removing the

effect/interaction did not affect the fit of the model (i.e., it is not a contributor to the fit of the model). As in Experiment 1, for all Bayesian analyses, only the BF_{10} is reported (i.e., the Bayes Factor in favour of the alternative hypothesis), and they are interpreted using the classification introduced by Schönbrodt and Wagenmakers (2017).

Second, to determine whether test performance improved as a function of the number of trained components, we ran a separate linear regression for each measure and dependent variable with Number of Components (0-3) as the predictor variable. Bayes factors were calculated for these analyses with the BayesFactor package using the default JZS prior. Significant regression models would therefore indicate that increasing the number of strategic task components relevant to the test included in the training tasks systematically affected test performance.

For the planned comparisons, instead of comparing performance from each training condition to all other conditions (as in Experiment 1), to reduce the number of comparisons, we limited analyses to comparisons between the *full schema* and *control* training groups to all other groups. Comparison of test performance in the *control* training group to the other groups made it possible to determine the extent to which training with any of the components relevant to the test task improved test performance relative to a condition in which the training was theoretically irrelevant to performing the test task. Comparison of test performance in the *full schema* training group to the other groups made it possible to determine the extent to which training with a subset of components relevant to the test task improved test performance relative to training with all components relevant to the test task.

Finally, in order to partition the cognitive processes involved in test performance into separable components, we fit the RT data from the test block to a Drift Diffusion Model[6]

[6] We thank Senne Braem for suggesting these analyses.

(DDM; Ratcliff, 1978) using the DstarM R package (van den Bergh et al., 2019). DDMs assume that participants performing binary decision tasks accumulate evidence until a response threshold is reached and can distinguish between four distinct parameters. Non-decision time (*t0*) reflects the time taken to complete low-level perceptual and/or motoric processes that are not involved in selecting an appropriate response. Boundary separation (*a*) can be considered a measure of the amount of information required to make a decision and is akin to a measure of participant cautiousness with regard to the task. A larger value in this parameter can be interpreted as indicating a more cautious approach to response selection (i.e., one resulting from the acquisition of more evidence before a decision is made). The starting point (*z*) is usually half way between the upper and lower boundaries (in the present study, these represent a correct or incorrect response respectively), but can also vary reflecting a degree of bias towards one response over the other. A larger value in this parameter would suggest some initial bias towards a correct response over an incorrect response. Finally, the drift rate (*v*) indicates the rate at which evidence is gathered and can be interpreted as reflecting processing efficiency. A larger value in this parameter indicates more efficient evidence accumulation toward a response decision.

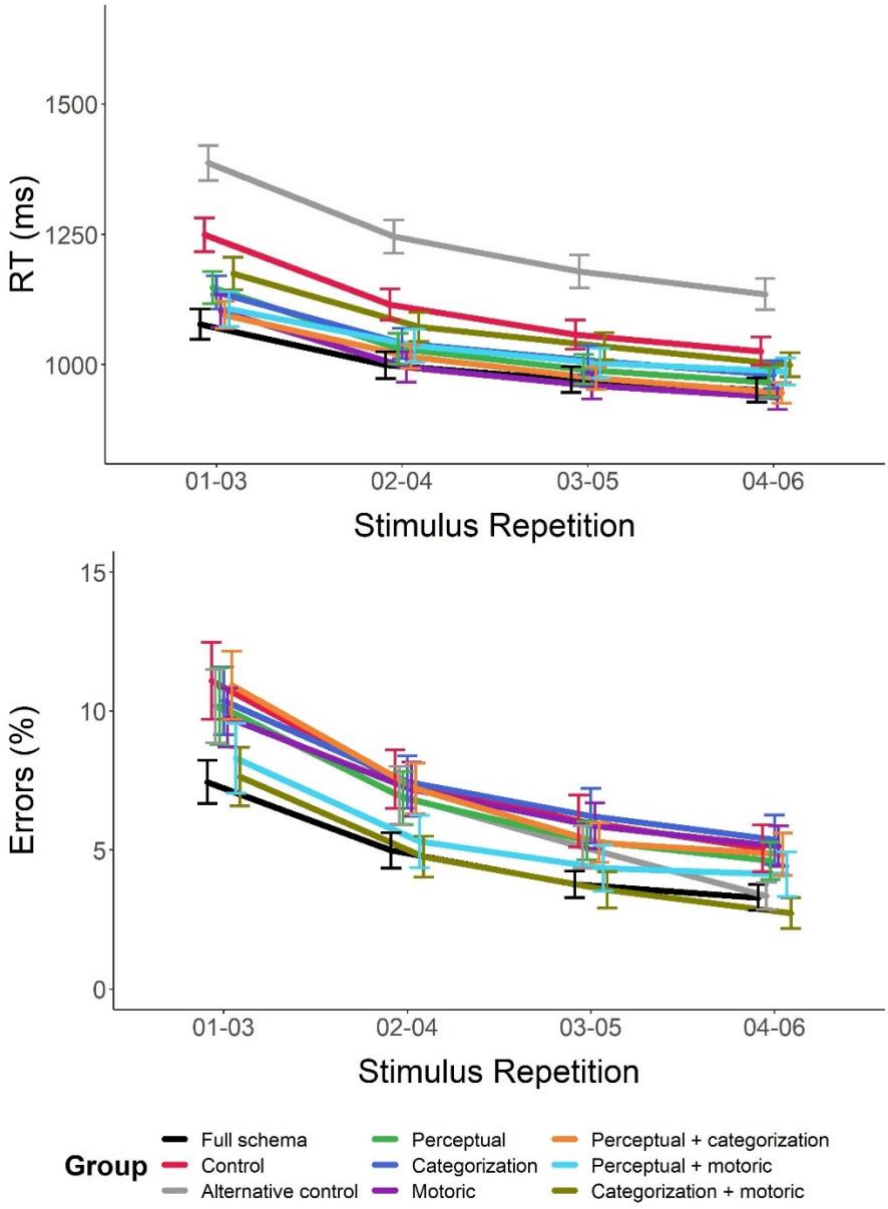
We fit the RT data from all valid test block trials (see data cleaning procedure above) to a DDM separately for each participant. Note that RTs from trials where a correct or an incorrect response was made were included in the DDM analysis. This strategy provides an overview of performance that is not limited to one measure or the other and instead looks at the balance between speed versus accuracy – an essential element of learning a novel task. In an attempt to reduce noise, we included all of the data from each test block rather than having distinct *initial* and *final* performance measures. There were no constraints set on any of the parameters. DstarM generated a value for each parameter based on test performance for every

participant. The resulting data were submitted to equivalent ANOVAs and regressions to those performed on the RT and error data using the same methods reported above.

Results and Discussion

Mean RTs and proportion of errors from the test block are plotted as a function of training group and stimulus repetition in Figure 4.

Figure 4
Mean RTs (top) and proportion of errors (bottom) from the test block of Experiment 2 plotted as a function of training group and stimulus repetition.



Note: Error bars show between-subjects difference adjusted 95% confidence intervals.

Omnibus analyses

The results from the between-subjects ANOVAs are reported in Table 3.

Table 3

ANOVA Results from Experiment 2. Equivalent Bayes Factors are also Reported.

	Initial performance				
	RT				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	7.43	0.007	0.012	0.316 ± 5.1%
Categorization	(1, 632)	1.88	0.171	0.003	4.279 ± 6.2%
Motoric	(1, 632)	3.76	0.053	0.006	2.074 ± 16.9%
Perceptual * Categorization	(1, 632)	0.25	0.620	<0.001	7.225 ± 8.2%
Perceptual * Motoric	(1, 632)	0.35	0.555	0.001	7.071 ± 9.2%
Categorization * Motoric	(1, 632)	5.59	0.018	0.009	0.546 ± 6.4%
Perceptual * Categorization * Motoric	(1, 632)	3.31	0.069	0.005	1.308 ± 7.3%
	Errors				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	0.36	0.547	0.001	9.436 ± 11.2%
Categorization	(1, 632)	0.78	0.377	0.001	8.473 ± 13.1%
Motoric	(1, 632)	7.92	0.005	0.012	0.223 ± 11.4%
Perceptual * Categorization	(1, 632)	0.67	0.414	0.001	6.268 ± 12.0%
Perceptual * Motoric	(1, 632)	0.15	0.697	<0.001	7.968 ± 12.8%
Categorization * Motoric	(1, 632)	0.82	0.367	0.001	5.677 ± 12.2%
Perceptual * Categorization * Motoric	(1, 632)	0.00	0.956	<0.001	5.787 ± 10.7%
	Final performance				
	RT				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	1.75	0.187	0.003	4.659 ± 12.3%
Categorization	(1, 632)	0.29	0.590	<0.001	9.863 ± 12.4%
Motoric	(1, 632)	0.37	0.543	0.001	10.009 ± 12.0%
Perceptual * Categorization	(1, 632)	1.08	0.300	0.002	6.862 ± 35.4%
Perceptual * Motoric	(1, 632)	1.84	0.175	0.003	4.154 ± 23.4%
Categorization * Motoric	(1, 632)	1.67	0.197	0.003	3.857 ± 12.1%
Perceptual * Categorization * Motoric	(1, 632)	3.01	0.083	0.005	1.392 ± 12.2%
	Errors				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	0.48	0.488	0.001	8.517 ± 5.3%
Categorization	(1, 632)	1.72	0.190	0.003	5.661 ± 12.1%
Motoric	(1, 632)	5.03	0.025	0.008	0.956 ± 6.5%
Perceptual * Categorization	(1, 632)	0.53	0.468	0.001	6.484 ± 8.9%
Perceptual * Motoric	(1, 632)	0.06	0.801	<0.001	10.798 ± 30.6%
Categorization * Motoric	(1, 632)	3.38	0.067	0.005	1.766 ± 5.4%
Perceptual * Categorization * Motoric	(1, 632)	0.64	0.423	0.001	4.954 ± 5.5%

Note: Bayes factors indicate whether removal of the effect/interaction from the full model would materially impair its fit. Thus, Bayes factors < 1 indicate that the effect/interaction is an important contributor to the model.

The ANOVA on initial RTs found that participants responded faster to the first three repetitions of each stimulus in the test block when the training included the same kinds of stimuli as the test task (mean RT = 1107 ± 15 ms) relative to when the training did not (mean RT = 1166 ± 16; main effect of Perceptual component: $p = 0.007$, $BF = 0.316 \pm 5.1\%$). Although there was a small benefit to initial RTs resulting from training with the kinds of

category templates (mean RT when the training included the same kinds of category templates = 1122 ± 15 ms; different category templates = 1152 ± 16 ms) and the kinds of responses (mean RT when the training included the same kinds of responses = 1116 ± 16 ms; different responses = 1158 ± 15 ms) used in the test task, neither of these differences reached significance (main effect of Categorization: $p = 0.171$, $BF = 4.279 \pm 6.2\%$; main effect of Motoric component: $p = 0.053$, $BF = 2.074 \pm 16.9\%$). However, the Categorization by Motoric components interaction was significant ($p = 0.018$, $BF = 0.564 \pm 6.4\%$) indicating that including the motoric component in the training regime was beneficial to test RTs when the training did not include the categorization component (mean difference = 144 ms). Whereas, when the training did include the categorization component, the benefit of including the motoric component was eliminated (mean difference = -9 ms).

None of the other effects or interactions from the ANOVA on initial RTs were significant (all *p*s > 0.06, all *BF*s > 1.3). None of the effects or interactions from the ANOVA on final RTs were significant (all *p*s > 0.08, all *BF*s > 1.3) indicating that any RT benefits of training were quite short lived and had been resolved by the end of the test block.

The ANOVA on the initial proportion of errors found that participants made fewer errors during the first three repetitions of each stimulus in the test block when the training included the same kinds of responses used in the test block (mean errors = 8.3 ± 0.5%) relative to when the training did not (mean errors = 10.6 ± 0.7%; main effect of Motoric component: $p = 0.005$, $BF = 0.223 \pm 11.4\%$)[7]. The ANOVA on the final proportion of errors found that this difference remained significant at the end of the test block (mean errors when the training included the same motoric component as the test task = 3.8 ± 0.3%; when the motoric component was different = 5.0 ± 0.4%; $p = 0.025$, $BF = 0.956 \pm 6.5\%$) indicating

[7] Note that the majority of errors were the correct code being entered for the incorrect category. Invalid codes (that did not identify either category) made up ~25% of all errors in the test block.

that the benefits of training with the motoric component of the test task had a degree of longevity. None of the other effects or interactions was significant in either ANOVA on response accuracy (initial errors: all *p*s > 0.3, all *BF*s > 5.6, final errors: all *p*s > 0.06, all *BF*s > 1.7).

The results from the regressions investigating the effect of number of relevant strategic task components included in the training tasks on test performance are reported in Table 4.

Table 4

Regression Results from Experiment 2. Equivalent Bayes Factors are also Reported.

	Intercept	DV	DF	*F*	*p*	*BF*	R^2	R^2 Adj
Initial RT (ms)	1,202.53	-43.87	(1, 638)	12.06	0.001	30.55	0.019	0.017
Initial errors (%)	11.26	-1.20	(1, 638)	6.18	0.013	1.784	0.010	0.008
Final RT (ms)	996.22	-14.51	(1, 638)	2.03	0.155	0.237	0.003	0.002
Final errors (%)	5.49	-0.73	(1, 638)	6.02	0.014	1.651	0.009	0.008

A significant regression equation was found for initial RTs ($p = 0.001$, $BF = 30.55$) indicating that participants' initial test RTs were 44 ms faster for each additional component included in the training. The regression on the final RTs was not significant suggesting that any benefits found at the start of the test block had been resolved by the end of the block.

A significant regression equation was found for initial errors and final errors (initial errors: $p = 0.013$, final errors: $p = 0.014$) suggesting some improvement in test response accuracy with each additional component included in the training. However, the Bayesian analyses found only anecdotal evidence in support of both models (initial errors: $BF = 1.8$, final errors $BF = 1.7$) suggesting that any effect of number of components on test accuracy was inconsistent at best.

Planned contrasts

The results from all planned contrasts comparing test performance across the training groups are reported in Table 5. For brevity, only the main findings of interest are summarized here.

Table 5

T-Test Results Comparing Test Performance Between Training Groups from Experiment 2. Equivalent Bayes Factors are also Reported.

	Initial performance							
	RT (ms)							
Contrast (training group)	**Difference**	**Lower *CI***	**Upper *CI***	***DF***	***t***	***p***	***BF***	q_{av}
Full schema vs. Perceptual	-70	-154	14	158	-1.66	0.100	0.601	0.26
Full schema vs. Categorization	-61	-146	24	158	-1.42	0.157	0.433	0.22
Full schema vs. Motoric	-25	-109	59	158	-0.58	0.560	0.200	0.09
Full schema vs. Perceptual + Categorization	-18	-94	59	158	-0.45	0.653	0.187	0.07
Full schema vs. Perceptual + Motoric	-29	-117	59	158	-0.65	0.518	0.207	0.10
Full schema vs. Categorization + Motoric	-97	-181	-13	158	-2.28	0.024	1.842	0.35
Full schema vs. Control	-171	-258	-85	158	-3.93	**<0.001**	167.409	0.61
Control vs. Perceptual	101	13	190	158	2.25	0.026	1.742	0.35
Control vs. Categorization	110	21	200	158	2.43	0.016	2.547	0.38
Control vs. Motoric	146	57	236	158	3.24	**0.001**	19.617	0.51
Control vs. Perceptual + Categorization	154	72	236	158	3.70	**<0.001**	78.240	0.58
Control vs. Perceptual + Motoric	142	50	235	158	3.04	**0.003**	11.103	0.47
Control vs. Categorization + Motoric	74	-15	163	158	1.65	0.102	0.593	0.25
	Errors (%)							
Contrast (training group)	**Difference**	**Lower CI**	**Upper CI**	**DF**	**t**	**p**	**BF**	q_{av}
Full schema vs. Perceptual	-2.7	-5.8	0.4	158	-1.74	0.084	0.681	0.27
Full schema vs. Categorization	-2.9	-5.8	-0.1	158	-2.03	0.044	1.122	0.31
Full schema vs. Motoric	-2.3	-4.9	0.3	158	-1.78	0.077	0.728	0.28
Full schema vs. Perceptual + Categorization	-3.5	-6.3	-0.6	158	-2.40	0.017	2.382	0.37
Full schema vs. Perceptual + Motoric	-0.9	-3.8	2.1	158	-0.58	0.564	0.199	0.09
Full schema vs. Categorization + Motoric	-0.2	-2.8	2.4	158	-0.15	0.883	0.172	0.02
Full schema vs. Control	-3.6	-6.8	-0.5	158	-2.29	0.023	1.889	0.36
Control vs. Perceptual	0.9	-2.9	4.8	158	0.47	0.642	0.189	0.07
Control vs. Categorization	0.7	-2.9	4.4	158	0.39	0.700	0.183	0.06
Control vs. Motoric	1.3	-2.1	4.8	158	0.76	0.450	0.222	0.11
Control vs. Perceptual + Categorization	0.2	-3.5	3.8	158	0.09	0.931	0.171	0.01
Control vs. Perceptual + Motoric	2.8	-0.9	6.5	158	1.49	0.139	0.471	0.23
Control vs. Categorization + Motoric	3.4	0	6.9	158	1.98	0.049	1.035	0.31
	Final performance							
	RT (ms)							
Contrast (training group)	**Difference**	**Lower CI**	**Upper CI**	**DF**	**t**	**p**	**BF**	q_{av}
Full schema vs. Perceptual	-15	-87	58	158	-0.40	0.692	0.183	0.06
Full schema vs. Categorization	-30	-100	41	158	-0.83	0.409	0.234	0.13
Full schema vs. Motoric	14	-52	80	158	0.41	0.679	0.185	0.06
Full schema vs. Perceptual + Categorization	5	-56	66	158	0.17	0.862	0.173	0.02
Full schema vs. Perceptual + Motoric	-36	-105	34	158	-1.02	0.311	0.274	0.16
Full schema vs. Categorization + Motoric	-48	-113	17	158	-1.47	0.144	0.460	0.23
Full schema vs. Control	-74	-145	-3	158	-2.07	0.040	1.206	0.32
Control vs. Perceptual	59	-18	137	158	1.52	0.131	0.494	0.23
Control vs. Categorization	45	-31	120	158	1.17	0.244	0.321	0.18
Control vs. Motoric	88	17	159	158	2.45	0.015	2.624	0.38
Control vs. Perceptual + Categorization	79	13	146	158	2.36	0.019	2.183	0.37
Control vs. Perceptual + Motoric	38	-36	113	158	1.03	0.307	0.277	0.16
Control vs. Categorization + Motoric	26	-44	96	158	0.73	0.467	0.218	0.11
	Errors (%)							
Contrast (training group)	**Difference**	**Lower CI**	**Upper CI**	**DF**	**t**	**p**	**BF**	q_{av}
Full schema vs. Perceptual	-1.3	-2.9	0.3	158	-1.61	0.109	0.562	0.25
Full schema vs. Categorization	-2.1	-4.1	-0.1	158	-2.07	0.040	1.216	0.32
Full schema vs. Motoric	-1.8	-3.5	-0.1	158	-2.14	0.034	1.390	0.33
Full schema vs. Perceptual + Categorization	-1.6	-3.3	0.2	158	-1.75	0.083	0.691	0.27
Full schema vs. Perceptual + Motoric	-0.8	-2.7	1	158	-0.89	0.372	0.247	0.14
Full schema vs. Categorization + Motoric	0.6	-0.9	2	158	0.77	0.442	0.224	0.12
Full schema vs. Control	-1.8	-3.7	0.1	158	-1.83	0.068	0.800	0.28
Control vs. Perceptual	0.4	-1.7	2.6	158	0.41	0.679	0.185	0.06
Control vs. Categorization	-0.3	-2.7	2.1	158	-0.25	0.801	0.176	0.04
Control vs. Motoric	-0.1	-2.3	2.1	158	-0.07	0.947	0.171	0.01
Control vs. Perceptual + Categorization	0.2	-2	2.5	158	0.19	0.849	0.173	0.03
Control vs. Perceptual + Motoric	0.9	-1.4	3.2	158	0.80	0.422	0.230	0.12
Control vs. Categorization + Motoric	2.3	0.3	4.3	158	2.31	0.022	1.961	0.36

Note: Unadjusted *p* values are reported, but only those printed in bold font remained significant following Holm-Bonferroni correction for multiple comparisons.

Initial test RTs were significantly faster in the *full schema* group relative to the *control* group ($p<0.001$, $BF=167.409$), indicating that training with all structural components relevant to the test task resulted in improved initial test performance relative to training with none of the relevant structural components of the test task. None of the other comparisons to the *full schema* group's initial RTs reached significance or survived correction for multiple comparisons. Taken together, these results suggest that, although training with the full schema expedited learning of the test task relative to the control, training with any of the structural task components relevant to the test task appeared to result in broadly equivalent benefits to initial test RTs. Nonetheless, it is important to compare performance from the active training groups to performance from the *control* group before drawing any strong conclusions regarding the benefits of training with a subset of strategic task components.

Initial RTs were significantly slower in the *control* condition relative to the *motoric*, *perceptual + categorization*, and *perceptual + motoric* groups (ps<0.004, BFs>11). Thus, in the present experiment, the training associated with the clearest RT benefits in the test task were the *full schema*, *motoric*, *perceptual + categorization*, and *perceptual + motoric* groups. That is, the greatest benefits to initial test RTs relative to the control condition were found when the training included practice with the full schema, only the kinds of responses, the kinds of stimuli and responses (presumably in part due to the benefits associated with practicing the motoric component), or the kinds of stimuli and categories used in the test task. The latter finding is especially noteworthy given that it suggests a benefit to test performance associated with performing the same kinds of dot-pattern classification tasks independently of training with the kinds of responses used at test.

None of the planned comparisons performed on the initial proportion of errors from the test block survived correction for multiple comparisons. Nor did any of the planned comparisons on the final performance measure for either RTs or errors. The latter findings

indicate that any performance differences between groups at the start of the test block had largely been resolved by the end of the test block suggesting that the benefits to initial test performance could be explained by task schema transfer and the benefits did not continue to expedite learning once the relevant task-set had been implemented.

Drift Diffusion Model

The results from the ANOVAs on each of the parameters from the DDM are reported in Table 6.

Table 6

ANOVA Results for Drift Diffusion Model Parameters from Experiment 2. Equivalent Bayes Factors are also Reported.

	t0				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	0.31	0.579	<0.001	10.284 ± 17.7%
Categorization	(1, 632)	0.51	0.478	0.001	8.054 ± 14.2%
Motoric	(1, 632)	0.65	0.419	0.001	7.876 ± 16.9%
Perceptual * Categorization	(1, 632)	0.33	0.569	0.001	6.188 ± 14.4%
Perceptual * Motoric	(1, 632)	0.28	0.600	<0.001	6.344 ± 14.2%
Categorization * Motoric	(1, 632)	2.93	0.088	0.005	1.834 ± 14.5%
Perceptual * Categorization * Motoric	(1, 632)	3.15	0.077	0.005	1.270 ± 14.4%

	a				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	4.42	0.036	0.007	1.461 ± 7.2%
Categorization	(1, 632)	2.60	0.108	0.004	3.072 ± 7.6%
Motoric	(1, 632)	0.26	0.610	<0.001	10.620 ± 7.1%
Perceptual * Categorization	(1, 632)	11.60	0.001	0.018	0.033 ± 10.1%
Perceptual * Motoric	(1, 632)	0.64	0.424	0.001	6.169 ± 9.7%
Categorization * Motoric	(1, 632)	3.10	0.079	0.005	1.930 ± 11.5%
Perceptual * Categorization * Motoric	(1, 632)	10.31	0.001	0.016	0.049 ± 8.3%

	z				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	1.77	0.184	0.003	4.322 ± 5.1%
Categorization	(1, 632)	0.06	0.811	<0.001	10.837 ± 8.7%
Motoric	(1, 632)	1.92	0.166	0.003	4.354 ± 4.8%
Perceptual * Categorization	(1, 632)	5.76	0.017	0.009	0.457 ± 5.4%
Perceptual * Motoric	(1, 632)	2.59	0.108	0.004	2.277 ± 7.5%
Categorization * Motoric	(1, 632)	0.26	0.614	<0.001	17.096 ± 59.8%
Perceptual * Categorization * Motoric	(1, 632)	2.07	0.151	0.003	2.218 ± 6.2%

	v				
Effect	**DF**	**F**	***p***	***η2***	***BF***
Perceptual	(1, 632)	1.11	0.293	0.002	6.170 ± 6.1%
Categorization	(1, 632)	0.14	0.705	<0.001	12.091 ± 5.9%
Motoric	(1, 632)	6.42	0.012	0.010	0.529 ± 6.0%
Perceptual * Categorization	(1, 632)	4.06	0.044	0.006	1.276 ± 6.0%
Perceptual * Motoric	(1, 632)	0.63	0.429	0.001	5.940 ± 6.3%
Categorization * Motoric	(1, 632)	0.10	0.748	<0.001	7.917 ± 8.0%
Perceptual * Categorization * Motoric	(1, 632)	4.06	0.044	0.006	0.951 ± 6.0%

Note: Bayes factors indicate whether removal of the effect/interaction from the full model would materially impair its fit. Thus, Bayes factors < 1 indicate that the effect/interaction is an important contributor to the model.

None of the effects or interactions from the ANOVA on the *t0* parameter reached significance (all *p*s > 0.07, *BF*s > 1.2) indicating that non-decision processes in the test task were broadly equivalent across all training groups.

The ANOVA on parameter *a* found a main effect of Perceptual training ($p = 0.036$) suggesting that those groups who gained some experience with the perceptual component during training (mean = 1.68) were less cautious than those who did not (mean = 1.73). However, the Bayesian analysis found anecdotal evidence that removal of this effect from the model would not materially impair its fit ($BF = 1.461 \pm 7.2\%$) suggesting that this effect should be interpreted with caution. Nonetheless, those participants who experienced training with both the perceptual and categorization components (mean = 1.62) were less cautious than participants who experienced neither the perceptual not categorization components (mean = 1.71), only the perceptual but not the categorization component (mean = 1.74), or only the categorization but not the perceptual component (mean = 1.75; Perceptual by Categorization interaction: $p = 0.001$, $BF = 0.033 \pm 10.1\%$). The 3-way interaction was also significant for parameter *a* ($p = 0.001$, $BF = 0.049 \pm 8.3\%$) indicating that the Perceptual by Categorization interaction was largely limited to those groups who also experienced training with the motoric component (mean values from smallest to largest: *Full schema* = 1.62, *Motoric* = 1.65, *Perceptual + Motoric* = 1.77, *Categorization + Motoric* = 1.80). Such an interaction was not present for the groups who did not experience training with the motoric component (mean values from smallest to largest: *Perceptual + Categorization* = 1.62, *Categorization* = 1.70, *Perceptual* = 1.71, *Control* = 1.77).

The ANOVA on parameter *z* also found a significant Perceptual by Categorization interaction ($p = 0.017$, $BF = 0.457 \pm 5.4\%$). Specifically, those participants who experienced training with both the perceptual and categorization components (mean = 0.63) started the trials with a larger bias towards a correct response than participants who experienced neither

the perceptual not categorization components (mean = 0.60), only the perceptual but not the categorization component (mean = 0.58), or only the categorization but not the perceptual component (mean = 0.54). Taken together with the results from the ANOVA on parameter *a*, this would suggest that training with both the perceptual and categorization components together produced the optimum conditions for rapidly learning the test tasks.

Finally, the ANOVA on parameter *v* found that participants processed the information more efficiently when their training did not include the motoric component (mean = 0.59) relative to when it did (mean = -0.09; main effect of Motoric training: $p = 0.012$, $BF = 0.529 \pm 6.0\%$). The Perceptual by Categorization interaction was also significant for this component ($p = 0.044$) indicating the most efficient processing of information in groups where the training included both perceptual and categorization components (mean = 0.72) relative to when the training included neither component (mean = 0.33), only the perceptual but not the categorization component (mean = 0.07), or only the categorization but not the perceptual component (mean = -0.11), However, the Bayesian analysis found anecdotal evidence that removal of the interaction would not materially impair the fit of the model ($BF = 1.276 \pm 6.0\%$) suggesting some degree of caution when interpreting this result .

To summarise, the findings from the ANOVA on the Drift Diffusion Model parameters suggest that non-decision time was largely unaffected by the differences in training between groups. This is especially noteworthy given that non-decision time is typically thought to represent low-level perceptual/motoric processes (van den Bergh et al., 2019), so one might expect to see some effect of perceptual and/or motoric training on this parameter. Nonetheless, as in the other analyses reported above, there was some evidence that training that included both the perceptual and categorization components, or just the motoric component alone was especially effective in reducing cautiousness (parameter *a*) and the speed at which information was accumulated (parameter *v*). This pattern of results seems to

suggest that training with the motoric component of the test task did not reduce non-decision time but did improve the two processes associated with cautiousness (parameter *a*) and efficiency in decision-making (parameter *v*). We return to this point in the General Discussion.

The results from the regressions investigating the effect of number of relevant training components on each of the parameters from the DDM are reported in Table 7. None of the parameters from the DDM was significantly affected by the number of components included in the training suggesting no systematic effect of increasing the number of trained components on any of the DDM parameters.

Table 7

Regression Results for Drift Diffusion Model Parameters from Experiment 2. Equivalent Bayes Factors are also Reported.

	Intercept	DV	DF	*F*	*p*	*BF*	R^2	R^2 Adj
t0	0.71	0.00	(1, 638)	0.31	0.579	0.102	0.000	-0.001
a	1.74	-0.02	(1, 638)	3.30	0.070	0.440	0.005	0.004
z	0.59	0.00	(1, 638)	0.03	0.866	0.089	0.000	-0.002
v	0.40	-0.10	(1, 638)	0.40	0.528	0.107	0.001	-0.001

General Discussion

The central aim of the present study was to further investigate the architecture of task schemas – learning strategies that can be applied to the early stages of engaging with a novel task. Two learning experiments were conducted to determine whether task schemas can be divided into separable components, each of which can be used to expedite learning a novel task that shares some of the structural components with the training task(s). To this end, different groups of participants received training with several unique tasks that included some/all/none of the structural components (the same kinds of stimuli, categories and/or responses), but none of the surface features (the specific stimuli, categories and/or responses) found in the test task (a dot-pattern classification task). The results from both experiments

indicated that task schema transfer was observable to a degree in all training conditions relative to the control group and that initial test performance improved as a function of the number of relevant components included in the training tasks. However, it was also clear that some components made a larger contribution to initial test performance than others (e.g., the motoric component alone, and the combination of perceptual and categorization components together). That is, the reported experiments found evidence that task schemas are componential and that practiced strategic task components can expedite learning novel tasks that share *some* of the structural elements represented in the schema.

Training with the motoric component of the schema was especially effective in the present experiments. One possible explanation for this could be in the extent to which the participants had made similar responses before. It is not clear whether the same pattern of results would have been observed with a more practiced (or simpler) motoric component such as a single key press. For example, the motoric aspects of driving a car are complex, but once they have been learned they can likely be applied to other similar contexts such as driving a bus, thereby expediting learning. Nonetheless, it is noteworthy that a substantial portion of the benefits to be gained by task schema transfer in the present experiments can be explained by a single (motoric) component of the schema. Future research would benefit from establishing the extent to which this was a specific feature of the present paradigm. However, the findings from Bhandari and Badre (2018), and Pereg et al. (2021), suggest that the benefits associated with task schema transfer are not limited to only the motoric component of the schema but can also be found in other components (e.g., the combination of perceptual and categorization training was also an important contributor to test performance in the present study; see below). Note that the required response in their experiments was a single key press that presumably did not require any additional practice, but they still found evidence of transfer of relevant strategies across unique tasks with a common structure.

Nonetheless, as with the present study, it is possible that a subset of the task schema components relevant to their tasks were especially important in improving performance, and that more directed training on those components alone would have also resulted in a similar pattern of results.

Similarly, in the present Experiment 2, there was evidence that learning about the perceptual and categorization components together (without the motoric component) readily transferred to the test task. A useful strategy in the present experiments might be to generate an internal representation of the category template from the pre-block instructions, and then superimpose that representation over the dot-patterns presented on each trial to determine their category membership. Such a strategy would be useful in the test task whatever action was required to indicate the relevant category. Note that training with a different kind of categorization task in the present Experiment 1 did not appear to improve participants' ability to perform the dot-pattern classification task at test (if anything, it seemed to hinder test performance). The latter observation is also consistent with the notion that training with task-specific strategic components can readily be applied to novel tasks with a common structure, but training with an irrelevant (or detrimental) strategy cannot (cf. Bhandari & Badre, 2018). As with the motoric component, further research would benefit from determining the extent to which this pattern of results was due to the features of the present test task where optimal performance relied on the ability to identify the category to which a dot-pattern stimulus belonged. Different test tasks might benefit from different training regimes depending on which components of the task schema were most useful in guiding performance.

The results from the Drift Diffusion Model analyses seem to suggest that training with the perceptual and motoric components of the test task did not affect non-decision time (typically thought to reflect low-level perceptual/motoric processes) but did affect participant cautiousness and the rate at which information relevant to the decision was accumulated. This

would imply that, at least in the current experiments, training of low-level processes did not affect the efficiency of these processes directly. Rather, training that included the perceptual and/or motoric strategic components of the test task appeared to affect task performance. This is consistent with the notion that participants were learning about the structure of the tasks and were able to apply that knowledge by rapidly developing novel learning strategies based on their prior experience. After all, the training did not include the specific stimuli/actions used in the test task but included only the *kinds of* stimuli/actions used at test. It is possible that training with specific stimuli and/or responses would have affected non-decision time if those same stimuli/responses were used at test by allowing participants to practice the specific perceptual and/or motor processes used in the test task thereby reducing non-decision time. Because novel tasks (with novel stimuli, categories, and responses) were introduced at test, non-decision time was not reduced by the training that was only able to support the learning of a new action rather than the application of a practiced action.

Training was effective to a degree in all groups at least in terms of response speed, despite the introduction of a task with altogether novel surface features at test. (Response accuracy at test was largely comparable across all groups and was typically quite high making it a less sensitive measure of the differences in performance between groups). There was also evidence that training with multiple strategic task components was more effective than training with fewer components suggesting an additive effect. The performance benefits associated with each training regime cannot be explained by familiarity with specific stimuli, categories, responses or combinations of these. Conversely, this pattern of results can easily be explained by the transfer of a task schema representing some/all of the critical strategic task elements (e.g., motoric, perceptual and/or cognitive control routines that are not bound to specific actions, percepts, or S-R bindings). Note that this view is also mirrored by the broad consensus in the transfer literature that (more abstract) conceptual relations readily transfer to

contexts that share an underlying structure (schema), but which differ on their surface features (e.g., Day & Goldstone, 2012).

These observations suggest that generating a task schema can accelerate acquisition of a novel task with a similar structure to the tasks practiced during training. Presumably, during the initial stages of the test phase, participants create a new task schema based on the instructions that they can use to formulate a new task-set over the course of the first few trials of the test block. This process may require a few trials of practice to stabilize (cf. Bhandari & Duncan, 2014; Brass, et al., 2017), or transition from relying on the task schema (the formulation of a learning strategy) to relying on a more stable task-set (i.e., transitioning from Chein & Schneider's, 2012, formation stage to their controlled execution stage). However, the evidence presented here is consistent with the notion that this process can be expedited by re-using *parts of* a recent task schema that shares some/all strategic components with the current task, even if the surface features of the task are novel and unique. The present work therefore extends Cole et al.'s (2010, 2011) theory of compositionality to include the recombining of strategic task components as well as specific task components. As noted above, the specific components that are most effective in improving test performance for a given task are likely determined by the features of the particular paradigm or test task.

For example, Pereg et al. (2021) observed transfer akin to that reported in the present work across multiple relatively simple two-choice reaction time tasks. Whereas Bhandari and Badre (2018) demonstrated the learning and transfer of relatively sophisticated control routines (WM input and output gating policies). In the present study, we examined control routines involved in relatively low-level (motoric/perceptual) mechanisms as well as more sophisticated (categorization) mechanisms and found evidence of both. Critically, the present experiments extend the findings of Pereg et al., and Bhandari and Badre by formalizing the componential nature of task schema architecture and also confirms what some theorists have

postulated in recent theories of learning and adaptive human behaviour – that components of individual learned routines at various levels of abstraction can be adapted to the current context thereby expediting learning (e.g., Cole et al., 2011; Taatgen, 2013; Verbruggen et al., 2014).

The present results are also consistent with recent theories of cognitive control (including computational models) which acknowledge that learning can include more abstract representations than specific S-R rules (e.g., Collins & Frank, 2013; Schmidt et al., 2016). For example, Collins and Frank (2013) have demonstrated that task-set structure can also be learned and applied to novel tasks that share a common structure. Their results even suggest that inferring a task-set structure is the norm and that these structures readily transfer between contexts even when they are detrimental to performance (see also: Dreisbach, 2012). The construction of task-sets (either via instruction or performance) has also received some attention recently in the instruction-learning literatures (for reviews, see e.g., Brass et al., 2017; Meiran et al., 2017). Although the details regarding this process and the mechanisms involved in the transfer of task schemas and their relationship with task-set construction remain murky, there is a growing recognition of the importance of such schemas in learning and adaptive human behaviour.

In the present experiments, the benefits of task schema transfer were relatively short lived – they were largely limited to the first few trials of the test block and were no longer observable by the end of the block. That is, the benefits of task schema transfer were largely limited to Chein and Schneider's (2012) formation stage of learning and had largely been eliminated by the time learning had transitioned to their controlled execution stage. This observation is consistent with the notion that task schema transfer was the source of the effect. As the new task becomes more practiced (and learning can rely more on the newly formed task-set), the benefits of task schema transfer dissipate. Apparently, the main benefit

to be gained by re-using components of recently formulated task schemas in the construction of novel schemas is in accelerating the formation of a novel task-set. Once that task-set has been created and has been applied to the current task a few times, the task schema itself is less active in controlling behaviour. Though apparently it is not altogether discarded because it can be used to expedite subsequent learning where appropriate.

It should be noted that this particular pattern of results is likely linked to the paradigm used here which was specifically designed to investigate transfer of task schemas[8]. Other forms of transfer certainly have longer-lasting effects on performance and are not limited to the first few attempts at learning a novel task. For example, individuals with more experience in reading experimental psychology journal articles will be able to decipher the content of this article more quickly than readers with less experience in this skill, and that benefit is likely to last beyond the first few words/sentences of the article. However, that class of transfer is more akin to Cole and colleagues' (2010, 2011) compositionality of learned instances – the recombining of existing representations/information in novel ways. Another example from a different domain of transfer would be that the transfer of category-response associations to novel stimuli from the same category (e.g., Longman et al., 2018) is likely to benefit performance for as long as the newly formed stimulus-category association remains intact. As discussed in the Introduction, transfer can take many forms, but the focus of the present study was on the transfer of (components of) learning strategies that can be used to accelerate the formation of novel task-sets. Nonetheless, it is possible that the rapid deterioration of transfer effects reported in the present experiments might be a feature of the paradigm used rather than our preferred interpretation.

[8] We thank an anonymous reviewer for pointing this out.

In the spirit of full disclosure, a third experiment was run using a similar paradigm to Experiment 2, but the results were largely inconclusive (i.e., almost all BFs from the planned contrasts comparing test performance from the *full schema* and *control* groups to the other groups were in the range 0.2-2.0). However, there was some evidence that the motoric component was also an especially important contributor to test performance in Experiment 3. This third experiment is described in more detail and the results are reported in the Supplementary Materials where we also report additional analyses that collapsed the data from Experiments 2 and 3 together. To summarise, the outcomes from the between-experiment analyses provided additional evidence that motoric training was especially important in directing test performance, and that test performance improved as a function of the number of relevant components included in the training tasks. The planned contrasts also added to the evidence that training with the perceptual and motoric components together or with the perceptual and categorization components together expedited the early stages of learning the test task.

There were some procedural differences between Experiments 2 and 3 (e.g., Experiment 3 had a longer practice phase), but it would be difficult to argue that any of these could explain the difference in outcomes between experiments. Although there were some qualitative differences in the patterns of results from Experiment 3, the most salient difference between the two experiments was in the variability of RTs and accuracy. Specifically, there was considerably more variability in the data from Experiment 3 than the data from Experiments 1 and 2. This might have reflected some differences in experience with participating in psychology experiments between the pools. For example, the participants from Experiments 1 (University of Exeter students) and 2 (recruited via Prolific) might have had more experience with participating in similar computer-based psychology experiments than the participants from Experiment 3 (students from the University of the

West of Scotland where participation in experiments is rare). However, we acknowledge that this is merely conjecture and that the source of the differences in variability of RTs/accuracy within the groups from each experiment remains unclear.

Another limitation of the present study is that learning difficulty was not matched across strategic task components in the present experiments[9]. It is possible that the performance benefits associated with training on components which were especially easy to learn might not be observable in test performance because asymptotic performance was reached almost immediately. Inspection of the learning data from the training phase of all three experiments (found in the Supplementary Materials) suggests that the perceptual, categorization, and motoric components (i.e., single-component training conditions) were the easiest to learn – performance reached asymptote before the end of the first training block and there was little additional improvement in performance throughout the remainder of the training phase. However, training with the motoric component was consistently associated with improved test performance in the present experiments suggesting that the ease by which this component was learned did not hide the transfer effect. Conversely, training with the full schema continued to improve performance throughout the practice phase suggesting that it was a difficult combination of components to learn. Nonetheless, full schema training was also one of the most reliably associated with improved test performance. Task/component difficulty was not explicitly controlled in the present experiments, and it would be difficult to equate the difficulty of learning a single component versus learning three. Nonetheless, the data seem to suggest that the relevance of each component to the test task was a better predictor of transfer than the speed at which each component was learned.

[9] We thank an anonymous reviewer for drawing this limitation to our attention.

To conclude, the current report makes a unique contribution by systematically examining the architecture and transfer of task schemas in a series of dot-pattern classification tasks. Specifically, by determining the kinds of information that can be represented therein, and how that information can be adapted to novel contexts. The findings suggest that task schemas can be divided into distinct components that can be reused to expedite the earliest stages of learning a novel task that shares some/all of the structural task components with the training task(s) and that this effect is additive. The particular components (or combinations of components) that are especially effective in optimising performance are likely to be determined by the format of the task and the prior experience of the learner. Nonetheless, the present study contributes to the growing literature investigating learning and transfer of task schemas by demonstrating that they can be divided into separable components which can individually be used to optimise learning a novel task.

References

Bhandari, A., & Badre, D. (2018). Learning and transfer of working memory gating policies. *Cognition*, 172, 89-100. DOI: 10.1016/j.cognition.2017.12.001

Bhandari, A., & Duncan, J. (2014). Goal neglect and knowledge chunking in in the construction of novel behavior. *Cognition*, 130, 11-30. DOI: 10.1016/j.cognition.2013.08.013

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.

Brass, M., Liefooghe, B., Braem, S., De Houwer, J. (2017). Following new task instructions: Evidence for a dissociation between knowing and doing. *Neuroscience & Biobehavioral Reviews*, 81, 16-28. DOI: 10.1016/j.neubiorev.2017.02.012

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural brain research*, 206 (2), 157-165. DOI: 10.1016/j.bbr.2009.08.031

Chein, J. M., & Schneider, W. (2012). The brain's learning and control architecture. *Current Directions in Psychological Science,* 21 (2), 78-84. DOI: 10.1177/0963721411434977

Cole, M. W., Bagic, A., Kass, R., & Schneider, W. (2010). Prefrontal dynamics in rapid instructed task learning reverse with practice. *The Journal of Neuroscience*, 30 (42), 14245-14254. DOI: 10.1523/JNEUROSCI.1662-10.2010

Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W., & Braver, T. S. (2011). Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Frontiers in human neuroscience, 5*. DOI: 10.3389/fnhum.2011.00142

Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering and generalizing task-set structure. *Psychological Review*, 120 (1), 190-229. DOI: 10.1037/a0030852

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, 47 (3), 153-176. DOI: 10.1080/00461520.2012.696438

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47 (1), 1-12. DOI:10.3758/s13428-014-0458-y

Dennis, I., & Perfect, T. J. (2013). Do stimulus-action associations contribute to repetition priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39 (1), 85-95. DOI: 10.1037/a0028479

Dreisbach, G. (2012). Mechanisms of cognitive control: The functional role of task rules. *Current Directions in Psychological Science,* 21 (4), 227-231. DOI: 10.1177/0963721412449830

Faul, F, Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. DOI: 10.3758/BF03193146

Hazeltine, E., & Schumacher, E. H. (2016). Understanding central processes: The case against simple stimulus-response associations and for complex task representation. In B. H. Ross (Ed.) *Psychology of Learning and Motivation* (pp. 195-245). New York: Academic Press. DOI: 10.1016/bs.plm.2015.09.006

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus-response bindings in priming. *Trends in Cognitive Sciences*, 18 (7), 376-384. DOI: 10.1016/j.tics.2014.03.004

Horner, A. J., & Henson, R. N. (2009). Bindings between stimuli and multiple response codes dominate long-lag repetition priming in speeded classification tasks. *Journal of experimental Psychology: Learning, Memory, and Cognition*, 35 (3), 757-779. DOI: 10.1037/a0015262

Horner, A. J., & Henson, R. N. (2011). Stimulus-response bindings code both abstract and specific representations of stimuli: evidence from a classification priming design that reverses multiple levels of response representation. *Memory and Cognition*, 39, 1457-1471. DOI: 10.3758/s13421-011-0118-8

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Phillip, A., & Koch, I. (2010). Control and interference in task switching - A review. *Psychological Bulletin*, 136, 849-874. DOI: 10.1037/a0019842

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4: 863, 1-12. DOI: 10.3389/fpsyg.2013.00863

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95 (4), 492-527. DOI: 10.1.1.114.7838

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, 22, 1-35. DOI: 10.1016/0010-0285(90)90002-L

Longman, C. S., Kiesel, A., & Verbruggen, F. (2020). Learning in the absence of overt practice: A novel (previously unseen) stimulus can trigger retrieval of an unpracticed response. *Psychological Research,* 84, 1065-1083. DOI: 10.1007/s00426-018-1106-4

Longman, C. S., Liefooghe, B., & Verbruggen, F. (2019). How does the (re)presentation of instructions influence their implementation? *Journal of Cognition,* 2 (1), 10. DOI: 10.5334/joc.63

Longman, C. S., Milton, F., Wills, A. J., & Verbruggen, F. (2018). Transfer of learned category-response associations is modulated by instruction. *Acta Psychologica,* 184, 144-167. DOI: 10.1016/j.actpsy.2017.04.004

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology,* 19 (2), 242-279. DOI: 10.1016/0010-0285(87)90012-0

Meiran, N., Liefooghe, B., & De Houwer, J. (2017). Powerful instructions: Automaticity without practice. *Current Directions in Psychological Science, 26*(6), 509–514. DOI: 10.1177/0963721417711638

Monsell, S. (2003). Task switching. *Trends in Cognitive Science*, 7 (3), 134-140. DOI: 10.1016/S1364-6613(03)00028-7

Morey, R. D., Rouder, J. N., & Jamil, T. (2022). *BayesFactor: Computation of Bayes factors for common designs (Version 0.9.12-4.3).*

Moutsopoulou, K., & Waszak, F. (2012). Across-task priming revisited: Response and task conflicts disentangled using ex-Gaussian distribution analysis. *Journal of Experimental Psychology: Human Perception & Performance,* 38 (2), 367-374. DOI: 10.1037/a0025858

Moutsopoulou, K., & Waszak, F. (2013). Durability of classification and action learning: Differences revealed using ex-Gaussian distribution analysis. *Experimental Brain Research,* 226 (3), 373-382. DOI: 10.1007/s00221-013-3445-0

Pashler, H., & Baylis, G. (1991). Procedural learning: 1. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17 (1), 20-32. DOI: 10.1037/0278-7393.17.1.20

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods,* 51, 195-203. DOI: 10.3758/s13428-018-01193-y

Pereg, M., Harpaz, D., Sabah, K., Ben-Shachar, M. S., Amir, I., Dreisbach, G., & Meiran, N. (2021). Learning the abstract general task structure in a rapidly changing task context. *Journal of Cognition*, 4 (1), 31. DOI: 10.5334/joc.176

Pfeuffer, C., Moutsopoulou, K., Pfister, R., Waszak, F. & Kiesel, A. (2017). The power of words: On item-specific stimulus- response associations formed in the absence of action. *Journal of Experimental Psychology: Human Perception and Performance,* 43(2), 328-347. DOI: 10.1037/xhp0000317

R Core Team (2022). *R: A language and environment for statistical computing (version 4.2.0)*. R Foundation for Statistical Computing, Vienna, Austria.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review,* 85 (2), 59-108. DOI: 10.1037/0033-295X.85.2.59

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The parallel episodic processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing costs. *Cognitive Psychology*, 91, 82-108. DOI: 10.1016/j.cogpsych.2016.10.004

Schumacher, E. H., & Hazletine, E. (2016). Hierarchical task representation: Task files and response selection. *Current Directions in Psychological Science,* 25 (6), 449-454. DOI: 10.1177/0963721416665085

Schneider, W., & Shiffrin, R. F. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 84 (1), 1-66. DOI: 10.1037/0033-295X.84.1.1

Schönbrodt, F. D., & Wagenmakers, E. J. (2017). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review.* Advance online publication. DOI: 10.3758/s13423-017-1230-y

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review,* 120 (3), 439-471. DOI: 10.1037/a0033138

Van den Bergh, D., Tuerlinckx, F., & Verdonck, S. (2019). DstarM: An R package for analysing two-choice reaction time data with the D*M method. *Behavior Research Methods,* 52, 521-543. DOI: 10.3758/s13428-019-01249-7

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, 136 (4), 601-626. DOI: 10.1037/a0019791

Verbruggen, F., McLaren, I., & Chambers, C. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives on Psychological Science,* 9 (5), 497-524. DOI: 10.1177/1745691614526414

Wickham H (2022). *ggplot2: Elegant Graphics for Data Analysis (version 3.3.5)*. Springer-Verlag New York.

Wills, A.J., Noury, M., Moberly, M.J., & Newport, M. (2006). Formation of category representations. *Memory and Cognition, 34*, 17-27. DOI: 10.3758/BF03193383

Contributions

Contributed to conception and design: CSL, FM, AJW

Contributed to acquisition of data: CSL

Contributed to analysis and interpretation of data: CSL, FM, AJW

Drafted and/or revised the article: CSL, FM, AJW

Approved the submitted version for publication: CSL, FM, AJW

Acknowledgements

We would like to thank Rosamund McLaren, Ruichong Shuai, and Sara Asanai for their help in data acquisition and Marius Golubickis for his help with the Drift Diffusion Model analyses.

Funding

This work was supported by the European Research Council (grant number 312445 awarded to Frederick Verbruggen), the Carnegie Trust for the Universities of Scotland (grant number RIG008617 awarded to Cai S. Longman) and the Experimental Psychology Society (Small Research Grant awarded to Cai S. Longman).

Conflict of interest

The authors declare that they have no conflict of interest.

Data accessibility statement

All raw data files, R scripts for data analysis, the scripts used for data collection, as well as the individual picture stimuli from the categorization + motoric training task of Experiment 1 are deposited on the Open Science Framework data repository (https://osf.io/t9whg/).