Learning in the absence of overt practice: A novel (previously unseen) stimulus can trigger retrieval of an unpracticed response

Cai S. Longman[1], Andrea Kiesel[2] and Frederick Verbruggen[3]

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

[2] University of Freiburg, Freiburg, Germany

[3] Ghent University, Ghent, Belgium

Abstract: 249 words. Text: 9924 words, 33 references, 4 figures, 3 tables.

Running head: Learning in the absence of practice

Corresponding author:
Cai S. Longman
Psychology
School of Media, Culture, and Society
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

This manuscript was accepted for publication in Psychological Research on 30/9/2018.

Abstract

Skilled performance is traditionally thought to develop via overt practice. Recent research has demonstrated that merely instructed stimulus-response (S-R) bindings can influence later performance and readily transfer across response modalities. In the present study, we extended this to include instructed category-response (C-R) associations. That is, we investigated whether merely instructed C-R bindings can trigger an unpracticed response (in a different modality) on perception of a novel (previously unseen) stimulus. In a learning-test design, participants had to classify stimuli by comparing them to perceptual category templates (Experiment 1) or semantic category descriptions (Experiment 2) presented prior to each block. During learning blocks, participants had to respond manually, respond vocally, or listen passively to the correct response being spoken. A manual response was always required at test. In test blocks, the categories could either be novel or repeated from the learning block, whereas half of the stimuli were always novel, and half were always repeated from the learning block. Because stimulus and category repetitions were manipulated orthogonally, it was possible to directly compare the relative contribution of S-R and C-R associations to performance. In Experiment 1, test performance was enhanced by repeating the C-R bindings independently of the stimulus. In Experiment 2, there was also evidence of an S-R repetition benefit independent of the classification. Critically, instructed associations formed in one response modality were robust to changes in the required response, even when no overt response was required during training, indicating the need to update the traditional view of associative learning.

Keywords: Instructed learning, associative learning, cognitive control, automatic retrieval, abstract representation.

Introduction

Theories of learning assume that skilled performance develops through the explicit pairing of a specific stimulus with a specific (overt) action. For example, Instance Theory (Logan, 1988, 1990) assumes that a task is initially performed on the basis of rules (algorithmic processing) but after every stimulus encounter a new processing episode is stored (for a similar proposal, see e.g., Schmidt, De Houwer, & Rothermund, 2016). When the stimulus is repeated, previous processing episodes are retrieved, facilitating performance when the stimulus-response (S-R) mapping is consistent and impairing performance when the mapping is inconsistent. Eventually, performance can rely entirely on memory retrieval and is said to be 'automatic' (see also: Chein & Schneider, 2012; Schneider & Shiffrin, 1977). Importantly, according to Instance Theory, "automaticity is specific to the stimulus and the situation experienced during training. Transfer to novel stimuli and situations should be poor" (Logan, 1988, p. 494). However, Logan also acknowledges that an instance might also include more abstract representations (e.g., categories) and suggests that such a definition of an instance might reconcile Instance Theory with evidence of transfer. The central aim of the current study was to determine the degree to which spontaneous transfer is found to occur between sets of stimuli and sets of responses (including an instructed response that does not require an overt action at all), with a particular emphasis on the response. That is, we were interested to determine the extent to which associations formed in one response modality would transfer to a different modality.

What is learned and how specific is it?

Most work on learning has focused on the formation of specific associations between observed stimuli and practiced responses (Hazeltine & Schumacher, 2016). However, recent research has elaborated the way a 'stimulus' and 'response' should be conceptualized (e.g., Dennis & Perfect, 2013; Henson, Eckstein, Waszak, Frings, & Horner, 2014; Horner & Henson, 2009,

2011) and has stressed the distinction of S-R associations from other kinds of associations (e.g., Hazeltine & Schumacher, 2016; Longman, Milton, Wills, & Verbruggen, 2018; Motsopoulou, Yang, Desantis, & Waszak, 2015; Moutsopoulou & Waszak, 2012, 2013). For example, Moutsopoulou and colleagues (2012, 2013, 2015) have compared the formation and durability of stimulus-action (S-A; the learned association between a specific stimulus and the specific action required for its classification) and stimulus-category (S-C; the learned association between a specific stimulus and its classification relevant to the current task) associations. On each trial in a prime/probe design, they asked participants to classify a picture of an everyday item according to a simple rule that was cued immediately before the target stimulus. Each prime phase (four trials) was immediately followed by a probe phase where every stimulus from the preceding prime phase was repeated, but the relevant classification (e.g., large/small, mechanical/non-mechanical) and/or the required action (e.g., left/right key press) could either switch or be repeated from that used during the prime phase. Over a series of experiments, Moutsopoulou and colleagues have consistently found evidence that S-A associations and S-C associations independently modulate performance when the classification and/or the required action change between prime and probe phases, indicating that (associative) learning need not be limited to S-R bindings alone (see also Dreisbach, 2012, for a review of recent research investigating the importance of task rules in modulating performance).

Longman, Milton, and colleagues (2018) have recently focused on the learning and transfer of S-R, S-C and category-response (C-R; the learned association between a category of stimuli and the relevant response) associations in a training-test design that used complex stimuli and responses. Participants were required to classify dot-patterns according to their overall similarity to two category templates presented prior to each block of trials by entering a complex (four-digit) response ‘code’ on the numeric keypad. Each training block used altogether novel stimuli, categories and responses, and was immediately followed by a test block in which the categories could either be novel or repeated from training. Critically, whether the test categories were novel or not, half of the stimuli to be classified at test were novel, and the remaining half were repeated from training (i.e., some of the dot-pattern stimuli could equally be classified according to two distinct categories; see Figure 1). Using complex responses made it possible to introduce altogether novel response codes for each training block, and to also manipulate response repetitions between training and test blocks orthogonally to stimulus/category repetitions.

By analyzing test performance from each condition, Longman, Milton, and colleagues (2018) could directly compare the relative contribution of three kinds of associations to learning. In test blocks where a repeated stimulus was classified using the same response as during training, but according to a novel category template, it was possible to determine the extent to which learned S-R associations transferred across classifications. In test blocks where a novel stimulus from a repeated category was classified by entering the same response code used during training, it was possible to determine the extent to which learned C-R associations transferred to novel stimuli from the same category. Finally, in test blocks where a repeated stimulus was classified according to the same category as during training by entering a novel response code, it was possible to determine the extent to which learned S-C associations transferred to novel responses (complex actions).

Longman, Milton, and colleagues found strong evidence that learned C-R associations rapidly transferred to novel stimuli from the same category, whereas the evidence for transfer of S-R associations across classifications and S-C associations to novel responses was much weaker: the improvement in test performance was much greater when the C-R bindings were repeated but a novel stimulus was classified, relative to when the S-R bindings were repeated but the classifications changed or when the S-C bindings repeated but the responses changed. This suggests that, at least in their complex perceptual classification task, learned C-R associations made a much greater contribution to test performance than did S-R or S-C associations (e.g., in their Experiment 4, η^2 from the analysis of RTs were: main effect of Category repetitions = 0.088; main effect of

Stimulus repetitions = 0.003; main effect of Response repetitions = 0.042). Furthermore, the findings indicate that (abstract) C-R associations readily transferred to novel stimuli. One of the aims of the present study was to examine transfer of the response as well.

It remains unclear whether the pattern of results reported by Longman, Milton and colleagues (2018) was limited to their artificial dot-patterns and perceptual categories, which might have encouraged participants to adopt a strategy that emphasized the rules rather than identification of specific exemplar stimuli. They avoided using familiar stimuli/categories because they were interested in the very early stages of learning and wanted to reduce the impact of prior learning on any transfer effects. For example, Kramer, Strayer, and Buckley (1990) noted that a part of transfer effects could be due to extra-category associations (e.g., 'cat' and 'dog' might be associated by the term 'raining cats and dogs') rather than the experimenter-defined category structure (e.g., 'pets'); it is also possible that any transfer effect would be limited to members of the category known prior to testing (e.g., 'cat' and 'dog' are well known pet animals, but an 'anole' (a green lizard) is a less common pet and might not benefit from the same transfer effect in an experiment that uses word stimuli). Nonetheless, it is possible that a different pattern of results might be found with pictures of everyday items and familiar semantic categories such as those used by Moutsopoulou and colleagues (2012, 2013, 2015). We tested this conjecture in the current Experiment 2.

It should be noted that the research of Longman, Milton and colleagues (2018) is broadly consisted with that of Pashler and Baylis (1991). The latter authors used well-learned semantic categories of familiar stimuli (e.g., numbers and letters) to investigate transfer of C-R associations to novel stimuli from the same category. In their experiments, participants had to classify stimuli by pressing one of three keys with their index, middle or ring finger. During the test phase, some of the training stimuli were repeated and some novel stimuli were also introduced. When conditions allowed it, they also found that learned C-R associations readily transferred to novel stimuli from the learned category (for a similar finding see also Kiesel, Wendt, & Peters, 2007). However, their design did not allow them to compare this effect to a 'pure' measure of S-R learning, independent of the classification (i.e., a condition where a given stimulus was classified using the same response, but according to a different category), nor could they rule out the possibility that their observed transfer effects were limited to members of the category that were known prior to testing (cf. Kramer et al., 1990). Note that both of these criticisms were addressed in the experiments of Longman, Milton et al. (2018), and in the current experiments (though the latter criticism is also true for the current Experiment 2).

Furthermore, Pashler and Bayliss used simple (single key press) responses thereby limiting the extent to which they could investigate transfer of C-R associations to novel responses. In one experiment they shuffled the relevant key press responses between training and test phases (Experiment 4) and found poor transfer, which could be explained by carryover of the (now incorrect) responses previously associated with a different category. In another experiment they switched the hand used to respond (Experiment 5). Although the relevant response changed (from the right to the left hand), transfer in the latter condition was relatively good, suggesting that transfer of responses is limited to actions semantically related to the training action with no prior (irrelevant) associations. However, it is less clear whether a trained response can also expedite learning of a semantically related response in a different modality, and/or whether an overt response is required at all for learning to take place. The current experiments were designed to clarify these points.

Learning via instruction

The research described above, primarily focused on learning via practice. More recently, several researchers have explored whether learning can develop in the absence of (overt) practice (e.g., Cohen-Kdoshay & Meiran, 2007, 2009; Liefooghe, De Houwer, & Wenke, 2013; Liefooghe, Wenke, & De Houwer, 2012; Longman, Liefooghe, & Verbruggen, 2018; Meiran, Cole, & Braver, 2012; Meiran, Pereg, Givon, Danieli, & Shahar, 2016; Meiran, Pereg, Kessler, Cole, & Braver, 2015a, 2015b; Verbruggen, McLaren, Pereg, Meiran, 2018). This rich line of research suggests that S-R associations can be formed via instructions alone. Importantly, these associations can 'automatically' influence performance in unrelated tasks (despite the lack of overt practice; *automatic effects of instructions*).

Inspired by the work of Moutsopoulou and colleagues (2012, 2013, 2015), Pfeuffer and colleagues (Pfeuffer, Hosp, Kimmig, Moutsopoulou, Waszak, & Kiesel, 2018; Pfeuffer, Moutsopoulou, Pfister, Waszak, & Kiesel, 2017; Pfeuffer, Moutsopoulou, Waszak, & Kiesel, 2018) extended instruction-based automaticity to S-A and S-C associations. Specifically, they adapted the priming paradigm of Moutsopoulou and colleagues such that, in some blocks (the 'verbal coding' blocks) participants did not have to perform any action during the priming phase but were merely required to passively listen to the classification and correct response being spoken to them via headphones (e.g., "mechanical, left") while the stimulus was visible[1]. Whether the response was 'verbally coded' or overtly executed during the prime phase, a manual response was required during the probe phase. Pfeuffer and colleagues found that classification and response repetitions, compared to switches, enhanced probe performance for a given stimulus whether an overt response was required during priming or not, though the effect was larger when a response was executed than when it was not (e.g., in Pfeuffer et al.'s (2017), Experiment 1, Cohen's *d* from the analysis of RTs were: executed response = 1.26; verbally coded response = 0.39). Critically, S-A and S-C associations were only repeated on 50% of probe trials so it would be counterproductive to

[1] Note that this design does not explicitly require participants to covertly practice the response, nor were participants explicitly instructed to withhold a response. Although there is a broader literature on mental practice (for reviews see: Driskell, Copper, & Moran, 1994; Schuster et al., 2011), this tends to focus on situations in which the participant is explicitly instructed to use mental imagery to practice the tasks. The current experiments, like those of Pfeuffer and colleagues' (2017, 2018), cannot rule out the possibility that participants used mental imagery to practice the tasks during learning blocks where a manual response was not required. However, our primary focus was whether the kind of learning generated in the absence of overt practice can transfer to an altogether novel context rather than the mechanism by which such associations might be formed.

learn/maintain them, thereby providing strong evidence that even incidental instructions can influence subsequent performance.

It remains to be seen whether the same pattern of results would be found in a more complex task – one that used more complex stimuli and responses, and where the transfer of learned C-R associations to novel stimuli (whether they were acquired via overt practice or not) is essential for optimum performance. Therefore, another aim of the current experiments was to extend the findings of Pfeuffer and colleagues (2017, 2018) to tasks that used more complex stimuli and responses – i.e., we wanted to test if C-R as well as S-R associations can be formed in the absence of an overt response[2], how readily these associations will transfer to novel stimuli/classifications, and whether they will transfer between response modalities[3].

The present study

To summarize, a central aim of the present study was to extend the findings of Longman, Milton et al. (2018) by examining the extent to which a complex response will transfer between response modalities. That is, we wanted to determine whether S-R and/or C-R associations formed in one response modality (e.g., a vocal response) would transfer to a different response modality (e.g., a manual response), thereby improving test performance even though the required response was inconsistent with the response performed during training (**Aim 1**). Furthermore, we also wanted to determine whether an abstract stimulus (category) representation can become associated with an abstract response representation (i.e., whether C-R associations can be formed) even when no action is required during training (**Aim 2**).

[2] Note that Cohen-Kdoshay and Meiran (2007, 2009) found a flanker compatibility effect on the first trials following some simple instructions that described the C-R bindings suggesting that C-R associations can be formed via instructions alone. However, the specific stimuli used in the subsequent block were also displayed during the instructions phase, so their findings could also be explained in terms of S-R bindings (which was indeed the preferred explanation of the authors).
[3] Note that Liefooghe et al. (2012) found evidence that S-R associations formed via instruction alone readily transferred between response modalities. However, some recent work from our own lab has found evidence to the contrary (Longman, Liefooghe, & Verbruggen, 2018).

A further aim of the current study was to extend the findings of Pfeuffer and colleagues (2017, 2018) by determining whether their results are limited to relatively simple left/right button presses which might have strong existing associations with their written/spoken equivalents. The latter conjecture might not be the case for more complex responses. Similarly, most instruction-based learning studies have used simple left/right responses. It is possible that overt practice in the relevant modality is a necessary condition for learning of complex responses. Furthermore, Pfeuffer and colleagues used the same (well-learned semantic) classifications and actions throughout their experiments. It is yet to be seen whether similar associations can be formed via instruction alone when the categories and responses (as well as the stimuli) are less familiar and change regularly. That is, we wanted to determine whether Pfeuffer and colleagues' observation that robust S-R associations can be formed in the absence of an overt response could be extended to tasks that use more complex stimuli, categories and responses (**Aim 3**).

Experiment 1

The current experiments were adapted from the design used by Longman, Milton, and colleagues (2018). Each experimental session was divided into pairs of 'learning' and 'test' blocks. Entirely novel stimuli, categories and response codes were introduced at the beginning of each learning block. Each block used eight different stimuli that were mapped to one of two possible categories (four stimuli per category). In Experiment 1, the stimuli were dot-patterns classified according to their overall similarity to (perceptual) category templates (like those used by Longman, Milton et al.). Participants were instructed that each category was mapped to one response code consisting of two-digit sequences. Each learning block was immediately followed by a test block in which the stimuli and/or categories could be repeated or not, while the response codes were always repeated between learning and test blocks. Importantly, stimulus and category repetitions were manipulated independently resulting in four critical conditions (see Figure 1). In one condition, the stimuli changed between learning and test blocks, but the categories and response codes were repeated (i.e., the test block used *Different stimuli*, but the *Same categories* and the *Same responses* as the preceding learning block: DsScSr[4]). In another condition the categories changed between learning and test blocks, but the same stimuli were presented, and they were mapped on the same response codes as those used during the preceding learning block (SsDcSr). In a third condition, the stimuli, categories, and response codes used at test were repeated from the preceding learning block (SsScSr). Finally, in the baseline condition, only the response codes were repeated between the learning and test blocks (DsDcSr). Experiment 1 also included three learning modality conditions: *Enter* (the participant had to enter the response code manually), *Speak* (the participant had to speak the response code vocally), and *Listen* (the participant had to listen to the correct response code being spoken to them via the computer speakers, equivalent to the 'verbal coding' condition of Pfeuffer et al., 2017, 2018).

Longman, Milton, and colleagues (2018) determined what was learned by computing the difference in performance between the end of the learning block and the start of the test block. In the present study, no performance data were collected for the learning blocks where no manual response was required (Speak and Listen). Therefore, we directly compared test performance across the different conditions. To examine specificity of response learning (Aim 1), we compared the DsScSr and SsDcSr conditions with the baseline condition in the Speak learning block: if test performance is worse in the baseline Speak condition than in the other two Speak conditions, transfer of learning must have occurred, despite the change in response modality between training (vocal) and test (manual). To examine if C-R and S-R learning and transfer could occur in the absence of an overt response (Aims 2 and 3, respectively), we examined transfer in the Listen

[4] The capital letters (and lower-case letters) indicate whether the stimuli (s), categories (c) and/or responses (r) of the transfer phase were the same (S) or different (D) from those used in the training phase. Although the responses were always repeated between training and transfer in the current experiment, we have used the same condition coding format as Longman, Milton and colleagues (2018). This was partly for consistency, but also to emphasise the relevant associations learned in each critical condition (SsDcSr = S-R association independent of the classification, DsScSr = C-R association independent of the stimulus).

conditions, using the same logic as for the Speak conditions. By comparing test performance in the SsDcSr and DsScSr conditions across the three learning modality conditions, we could also further contrast S-R and C-R learning, like Longman, Milton, and colleagues.

Method

Participants. 60 students from the University of Exeter (47 Female) with a mean age of 20.2 years (*SD*=2.5) were paid £7 or awarded partial course credits for their participation. Prior to data collection we set the target sample size to 42 (with N = 42, it was possible to detect medium-sized differences). However, the pattern of results was not conclusive, so we tested a further 18. To correct for this (single) 'peek', we set the alpha level to .025 in order to reduce the chances of making a Type I error (i.e., accounting for the fact that we had performed an extra analysis; cf. Strube, 2006). Exclusion criteria (see below) were set in advance of data collection. 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.

Apparatus, stimuli, and responses. Stimuli were presented on a 21.5-inch iMac using Psychtoolbox (Brainard, 1997). As shown in Figure 2, the stimuli 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 = 0.1 cm). Randomizing the location of each stimulus ensured that the stimuli could not be classified according to the absolute location of some/all of the dots, and therefore encouraged comparison with the category templates. In each block, there were two categories (selected pseudo-randomly), and eight stimuli (four exemplar stimuli per category). All category templates and their labels are shown in the Appendix. Of the five dots in each exemplar stimulus, three were positioned within the borders of the category template and the remaining two were presented at adjacent locations (note that the category templates were only presented during the pre-block instructions and were not displayed during the trials; see Figure 1 for some examples). Thus, the category membership of a given dot-pattern was determined by its overall similarity to the category templates presented at the start of each block. The relative location of each dot in a given stimulus was determined 'by hand' (i.e., not randomly selected) to ensure that some stimuli could equally be classified according to two distinct category templates. Likewise, one set of category templates (and stimuli) were used during learning blocks, whereas the additional templates (and stimuli) were only introduced at test (where necessary; see Appendix 1). Naturally, some stimuli were better matched to the template(s) than others, but the specific categories used in each experimental condition was randomized anew for each participant; therefore it seems unlikely that this could have influenced the results reported below.

The primary motivation behind using complex two-digit response codes was to ensure that novel responses could be associated with each new category pair introduced throughout the experiment. The two digits in each response code were always on adjacent keys in order to equate the difficulty of entering each code. Two different codes (one per category) were used for each learning-test block pair. Selection of the codes used for each pair was pseudo-randomized to discourage re-classification according to a simple spatial rule and to equate the difficulty of the tasks (e.g., when used in the same block, the codes 41 and 47 could easily be reclassified as 'up' and 'down' respectively because the only difference between them is in the second digit – either an 'upward' or 'downward' motion with the index finger; therefore, 41 vs. 47 might be easier to memorize than 23 vs. 47).

There were three learning modality conditions (Enter, Speak, Listen) manipulated within subjects. In all test blocks participants had to indicate to which category the stimulus belonged by entering the relevant two-digit response code (i.e., a manual response was always required at test). In all blocks (learning and test) for all conditions, once the response had been entered, spoken or heard, participants had to press the space bar with their left index finger to move on to the next trial. This allowed each condition to follow a standardized procedure.

Procedure. At the beginning of the experiment, participants were informed that each block would start with a pre-block instructions screen displaying the two category templates, the category names and relevant response codes used in the subsequent block for 15 seconds. This was followed by a screen displaying the response modality required in the subsequent block for 3 seconds (see Figure 2). Each block ended with a feedback screen displaying the number of errors made during the ‘catch trials’ for 15 seconds (procedure detailed below). Although none of the response codes included the number 5, in blocks that required a manual response, participants were instructed to place their right index finger on the (central) 5 key in the numeric keypad. They were also instructed to place their left index finger on the space bar and because, on some (catch) trials, they would be required to respond using the ‘F’ or ‘V’ keys (see below), they should place their left-hand ring and middle fingers on these keys from the outset. They were also instructed to respond as quickly as possible while minimizing errors.

The procedure for the learning blocks differed depending on the response condition. In all conditions, each trial started with a blank screen presented for 500 ms, followed by the target stimulus, which remained visible for 1000 ms. The screen then remained blank until the space bar was pressed to indicate the end of the trial. A reminder screen was then presented for 1000 ms: the category name and the correct response code were centrally presented in black Ariel font (size 30, see Figure 2). Note that this reminder screen did not explicitly specify whether the executed response was correct or not. Prior to pressing the space bar to end each trial, participants had to either classify the stimulus by entering (Enter) or saying (Speak) the correct two-digit code for the relevant category or they had to passively listen to the correct code being spoken to them (Listen) via the computer speakers without making an overt response. In the latter condition, the entire response code took 800 ms to play and started 100 ms after stimulus onset. In order to ensure participants did not practice the overt manual response in the Speak and Listen conditions (e.g., by manually entering the response code) they were asked to keep a clenched fist with their right hand next to the keyboard (i.e., visible to the experimenter). During the test block of all conditions, the procedure was identical to the procedure for learning blocks in the Enter condition (see Figure 2).

To encourage participants to pay attention to the vocal codes in the Listen condition, we introduced four (pseudo-randomly selected) ‘catch’ trials during the learning blocks for all conditions (two catch trials were included in the test blocks). On these trials, immediately following the space bar response, the participant was presented with two response codes (above and below the centre of the screen): the correct code for the trial, and the incorrect code (i.e., the other code used as a response during the block). The participant was required to indicate which code was the correct response by pressing either ‘F’ (upper code) or ‘V’ (lower code) with their left ring or middle finger respectively.

Learning and test blocks alternated throughout the experiment (i.e., learning-test-learning-test-…). In each block, eight stimuli were used that belonged to one of the two instructed categories. Each of the eight stimuli was presented five times during learning blocks (resulting in 40 learning trials) and three times during test blocks (resulting in 24 test trials). Novel stimuli, categories and responses were introduced for each learning block. In half of the test blocks, the categories used in the preceding learning block were repeated; in the other half, the categories changed. Whether the categories repeated or not, half of the stimuli used during test blocks were the same as those used in the preceding learning block and the remaining stimuli were novel (see Figure 1 for some examples). The response codes were always repeated between learning and test blocks. This resulted in four transfer conditions equivalent to those used by Longman Milton et al. (2018): DsDcSr, SsDcSr, DsScSr, SsScSr. Note that each test block consisted of two transfer conditions: test blocks in which the categories changed consisted of DsDcSr and SsDcSr trials; whereas test blocks in which the categories were repeated consisted of DsScSr and SsScSr trials (see Figure 1).

Whether the category templates used at test were the same or different to those used during learning blocks alternated through the experiment and the order was counterbalanced over participants. Both pairs of transfer conditions were completed for a single learning response modality (e.g., Speak) before moving on to the next modality (order counterbalanced over participants). Because each transfer condition was performed in each learning response modality, there were 12 conditions in total (e.g., Enter DsDcSr, Speak SsDcSr, Listen DsScSr, …).

The experiment started with a short practice phase consisting of two 48-trial blocks. Note that we did not show any dot patterns in the practice phase; thus, participants only practiced entering/saying response codes. During practice, each trial started with a blank screen presented for 500 ms followed by a response code displayed centrally (black Ariel font, size 30) for 1000 ms. The screen then remained blank until the two-digit response had been entered and the space bar had been pressed. On each trial, the participant was required to enter the code on the numeric keypad using their right index finger and also speak the code vocally. Immediate feedback (given on every trial) was visible for 1000 ms and displayed the correct response code in either green (correct) or red (error). Average correct RT and response accuracy were presented at the end of each practice block for 15 seconds. During practice, every possible response code (24 in total) was presented four times (twice per block). Data from the practice phase was not analyzed.

The experimental phase consisted of twelve consecutive learning-test block pairs (four per learning response modality condition: 2 learning-test block pairs where novel categories were introduced at test and 2 where the categories were repeated) with a timed 2-minute break after the sixth learning-test block pair (i.e., half way through the experimental phase). Performance from the stimuli that belonged to the same transfer condition were pooled. This resulted in a total of 24 test trials being available for analysis per condition. The experiment consisted of 768 experimental trials and an additional 96 practice trials and lasted approximately one hour.

Analyses. All data processing and analyses were performed using R (R Development Core Team, 2018). Raw data files and R scripts from all experiments (as well as the Matlab scripts and audio files from all experiments and the picture stimuli used in Experiment 2) are deposited on the Open Science Framework data repository (https://osf.io/uw2bm/).

Performance from only the test block in each condition was analyzed. Trials with RT (first digit) < 100 ms (0.01%) and trials with RT (second digit) > 4000 ms (0.96%) were omitted from all analyses. Error trials were omitted from the RT analyses. The data from twelve participants (10 participants from the original 42, plus 2 additional participants from the subsequent 18) were replaced because they had <50% of the maximum possible observations in at least one condition following the above data cleaning procedures indicating the difficulty of the task[5]. The latency of the second digit in the response code (i.e., the time it took to enter the entire code) was used for all response latency analyses.

For each dependent variable (RT and proportion of errors), we performed a separate Modality (Enter, Speak, Listen) by Stimulus (same, different) by Category (same, different) ANOVA. We also report Bayes factors and effect sizes (generalized eta squared) 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, interactions were only allowed if all constituent sub-effects were also included (see Morey et al., 2015). When this approach is used, Bayes factors <1 indicate that removing the effect/interaction from the full model is deleterious (i.e., is a contributor to the fit of the full model).

[5] In order to confirm that replacing these participants did not materially affect the pattern of results, we performed the omnibus ANOVAs described below including all 72 participants. The pattern of results was almost identical – strong evidence of C-R transfer, little evidence of S-R transfer, no significant interactions with the Modality factor.

Results and Discussion

The RTs and proportion of errors made in the test blocks are plotted as a function of stimulus and category repetitions, separately for each learning response modality in Figure 3. The results from the omnibus ANOVAs are reported in Table 1.

As predicted, the omnibus ANOVAs showed that test performance was better when the categories were repeated from training (mean RT = 1297 ms, errors = 12.0%) relative to when they were novel (mean RT = 1385 ms, errors = 17.9%; main effect of Category for RTs: $p<.001$, $BF<0.001 \pm 28.1\%$; main effect of Category for errors: $p<.001$, $BF<0.001 \pm 3.9\%$) indicating that repeating the C-R bindings improved test performance. Absolute test performance was also modulated by the response modality used during learning blocks with the best test performance found when participants had to enter the response code manually during learning blocks (Enter RT = 1307 ms, errors = 13.2%), intermediate when they had to speak the response code (Speak RT = 1349 ms, errors = 15.5%), and worst when they merely had to listen to the response code (Listen RT = 1368 ms, errors = 16.1%); main effect of Modality for RTs: $p=.028$, $BF=0.019 \pm 5.1\%$[6]; main effect of Modality for errors: $p=.018$, $BF=0.259 \pm 8.0\%$. The main effect of Stimulus failed to reach significance for both RTs ($p=.818$, $BF = 12.16 \pm 4.4\%$) and errors ($p=.048$, $BF = 3.49 \pm 13.8\%$) indicating that test performance was largely unaffected by repetitions of the S-R bindings, whether the classifications were repeated from the preceding learning block or not.

Figure 3 shows an interaction between stimulus and category repetition. When novel categories were introduced at test, test RTs were slower for those stimuli that had previously been classified under a different category (SsDcSr RT = 1400 ms) relative to altogether novel stimuli (DsDcSr RT = 1371 ms); by contrast, when the categories were repeated between the learning and test blocks, test RTs were faster for familiar stimuli (SsScSr RT = 1281 ms) relative to novel stimuli

[6] Note that $p>.025$ so it just failed to reach significance according to the adjusted alpha. However, the Bayesian analysis (which does not require adjustment) provided very strong evidence that removal of the main effect of Modality from the model would impair its fit.

from the same category (DsScSr RT = 1314 ms). This pattern of results suggests that a familiar stimulus automatically retrieves S-C associations formed during the learning phase which was beneficial to test performance when the classification was repeated but was detrimental to performance when the classification changed. The Stimulus by Category interaction was reliable for RTs ($p<.001$; though the Bayesian analysis found only anecdotal evidence in support of the interaction, $BF=0.451 \pm 5.6\%$) but not errors ($p=.541$, $BF=9.04 \pm 16.0\%$).

Importantly, none of the interactions involving the Modality factor were reliable for RTs (ps>.33, BFs>8.8) or errors (ps>.03, BFs>0.5[7]). Taken together with the Bayesian analyses and the other effects/interactions from the omnibus ANOVA, the latter result provides evidence that C-R associations readily transfer to both old and novel stimuli from the same category, whether the response had been overtly practiced (Enter), overtly practiced in a different response modality (Speak) or had not been overtly practice at all (Listen). Thus, Experiment 1 suggests that learning is not action-specific (Aim 1) and can even develop without experience of the stimulus or the action (Aim 2). The current experiment also replicated the observed bias toward learning the C-R bindings independently of the stimulus over learning the S-R bindings independently of the classification reported by Longman, Milton and colleagues (2018). However, we found little evidence that learned S-R associations transferred across classifications whether the response had been overtly practiced (Enter), or not (Listen). That is, we were not able to replicate the findings of Pfeuffer and colleagues (2017, 2018) in an experiment that used a complex task (Aim 3). The question remains whether this pattern of results is specific to perceptual classification tasks. We addressed this issue in Experiment 2.

[7] Note that the Modality by Category interaction just failed to reach significance according to the adjusted alpha. However, the Bayesian analysis (which does not require adjustment) provided only anecdotal evidence that removal of the interaction from the model would impair its fit. Inspection of Figure 3 suggests that the accuracy advantage found when the categories were repeated between training and transfer was smallest in the Enter condition, intermediate in the Speak condition and largest in the Listen condition. However, these differences were apparently small and unstable.

Experiment 2

The stimuli and categories used in Experiment 1 (and in the experiments of Longman, Milton, et al., 2018) were artificial dot patterns classified according to perceptual category templates and were neither typical for experiments that investigate basic instructed/associative learning mechanisms (e.g., Horner & Henson, 2009, 2011; Moutsopoulou et al., 2012, 2013, 2015; Pfeufer et al., 2017, 2018), nor familiar to everyday life. This feature of the design was intentional – we wanted to minimize the effects of prior experience (see Kramer et al., 1990, for a discussion of prior experience in learning experiments). However, it is possible that this influenced the pattern of results. Experiment 2 was designed to test this conjecture.

To this end, we repeated the procedure for Experiment 1, but replaced the dot patterns with drawings of everyday items classified according to familiar semantic categories. We therefore tried to replicate our findings from Experiment 1 with a design that is more akin to other research investigating instructed and/or practice-based learning while also improving the ecological validity of our original study. If the bias toward learning the C-R associations (independently of the stimulus), over learning the S-R associations (independently of the classification) as observed in the current Experiment 1 and Longman, Milton and colleagues (2018) is also observed in the current Experiment 2, then we can conclude that the effect is not limited to paradigms that use complex stimuli and perceptual categories.

Method

60[8] different students from the University of Freiburg (41 Female) with a mean age of 27.0 years (*SD*=8.6) were paid €10 or awarded partial course credits for their participation. We used the same exclusion criteria as Experiment 1, but it was not necessary to replace any participants

[8] Note that the target sample size was 60 from the outset in Experiment 2 so there was no need to adjust the alpha.

following data cleaning (see below). Written informed consent was obtained after the nature and possible consequences of the study were explained.

The design and procedure of Experiment 2 was identical to Experiment 1 apart from the following critical difference: the stimuli were black drawings of everyday items (up to 5 cm wide and 5 cm tall, displayed centrally on a white background) that were classified according to familiar semantic categories. As in Experiment 1, some stimuli could equally be classified as being members of two categories (all categories are listed in the Appendix). For example, a carrot could be classified as a vegetable or a food that is crunchy, whereas a banana could be classified as a fruit or a food that is soft. The pre-block instructions screen was slightly different to Experiment 1 because there were no category templates to display. Instead participants were presented with the category labels and the correct response for each category as well as a short description of each category (arranged as in Experiment 1; see Figure 2).

Testing for Experiment 2 was conducted in a different lab to Experiment 1. The stimuli were presented on a 24-inch LED monitor on a Fujitsu Eprimo P920 computer running Matlab. Because the lab was in a German university, all of the instructions (including category names/descriptions) and the voice recordings of the numbers used in the Listen condition were translated into German.

As in Experiment 1, trials with RT (first digit) < 100 ms (0.05%) and trials with RT (second digit) > 4000 ms (0.82%) were omitted from all analyses. Error trials were also omitted from the RT analyses. We performed identical analyses to those performed in Experiment 1, except we also ran Stimulus (same, different) by Category (same, different) ANOVAs on each dependent variable, separately for each Modality condition in order to unpack the interactions involving the Modality factor in the omnibus ANOVAs (note that these additional analyses were not necessary in Experiment 1 where there were no significant interactions involving the Modality factor in the omnibus ANOVAs). Equivalent Bayes factors and effect sizes are also reported.

Results and Discussion

The RTs and proportion of errors made at test are plotted as a function of stimulus and category repetitions, separately for each learning response modality in Figure 4. The results from the omnibus ANOVAs are reported in Table 2 and the results from the follow-up ANOVAs are reported in Table 3.

As predicted, the omnibus ANOVAs showed that test performance was better when the categories were repeated from the preceding learning block (mean RT = 1080 ms, errors = 5.5%) relative to when they were novel (mean RT = 1144 ms, errors = 7.3%; main effect of Category for RTs: p<.001, BF<0.001 ± 11.9%; main effect of Category for errors: p=.001, BF=0.003 ± 6.7%) indicating that, as in Experiment 1, repeating the C-R bindings improved test performance. However, unlike Experiment 1, performance was also better when the stimuli were repeated from the preceding learning block (mean RT = 1090 ms, errors = 5.8%) relative to when they were novel (mean RT = 1134 ms, errors = 7.0%; main effect of Stimulus for RTs: p<.001, BF<0.001 ± 11.9%; main effect of Stimulus for errors: p=.002, BF=0.284 ± 7.3%) indicating that repeating the S-R bindings also improved test performance. That is, the bias toward learning the C-R bindings independently of the stimulus over learning the S-R associations independently of the classification observed in the experiments of Longman, Milton and colleagues (2018) and Experiment 1 of the present study was not replicated in the current experiment that used easily identifiable picture stimuli classified according to familiar sematic categories. As in Experiment 1, absolute test performance was also modulated by the response modality used during learning blocks, with the best test performance found when participants had to enter the response code manually during learning blocks (Enter RT = 1070 ms, errors = 4.9%), intermediate when they had to speak the response code (Speak RT = 1125 ms, errors = 6.7%), and worst when they merely had to listen to the response code (Listen RT = 1141 ms, errors = 7.7%; main effect of Modality for RTs: p<.001, *BF*<0.001 ± 12.3%; main effect of Modality for errors: *p*<.001, *BF*<0.001 ± 6.2%) indicating that test performance was improved when the training included a manual response.

Figure 4 shows an interaction between stimulus and category repetition, but its pattern is slightly different to Experiment 1. Whether the categories used at test were repeated from the previous learning block or not, test performance was better for those stimuli that had been previously classified (SsDcSr RT = 1131 ms, errors = 6.8%; SsScSr RT = 1049 ms, errors = 4.8%) relative to novel stimuli (DsDcSr RT = 1157 ms, errors = 7.9%; DsScSr RT = 1111 ms, errors = 6.2%). However, this difference was smaller for novel categories (DsDcSr-SsDcSr difference: RTs = 47 ms, errors = 1.7%) relative to when the categories were repeated from the preceding learning block (DsScSr-SsScSr difference: RTs = 82 ms, errors = 2.0%). The Stimulus by Category interaction was reliable for RTs (*p*=.003; though the Bayesian analysis found only anecdotal evidence that removal of the interaction from the model would impair its fit, *BF*=0.521 ± 14.5%) but not errors (*p*=.777, *BF*=8.854 ± 11.4%). The Modality by Stimulus interaction was also reliable for RTs (*p*<.001, *BF*=0.065 ± 12.1%) but not errors (*p*=.059, *BF*=3.947 ± 6.0%) suggesting some modulation of the size of the stimulus repetition benefit by the learning response modality. However, the Modality by Category interaction did not approach significance for either dependent variable (*p*s>.10, *BF*s>4.5) suggesting that the category repetition benefit was not modulated by the response modality used during learning blocks. The three-way interaction did not approach significance in either dependent variable (*p*s>.20, *BF*s>8.1).

The ANOVAs performed separately for each learning response modality condition all found that test performance was better when the categories were repeated from the preceding learning block (Enter: mean RT = 1038 ms, errors = 4.1%; Speak: mean RT = 1084 ms, errors = 5.8%; Listen: mean RT = 1118 ms, errors = 6.6%) relative to when they were novel (Enter: mean RT = 1102 ms, errors = 5.7%; Speak: mean RT = 1166 ms, errors = 7.6%; Listen: mean RT = 1164 ms, errors = 8.8%). This was confirmed by a reliable main effect of Category for both dependent variables in all training modalities (Enter RTs: *p*<.001, *BF*<0.001 ± 3.6%; Enter errors: *p*=.034, *BF*=0.350 ± 5.6%; Speak RTs: *p*<.001, *BF*<0.001 ± 7.2%; Speak errors: *p*=.013, *BF*=0.224 ± 5.1%; Listen RTs: *p*<.001, *BF*=0.001 ± 5.0%; Listen errors: *p*=.048, *BF*=0.329 ± 6.1%). Taken alongside the lack of any reliable Modality by Category interactions in the omnibus ANOVAs (and the Bayesian analyses which found substantial evidence that removal of the interaction from the model would not materially impair its fit), the evidence would suggest that transfer of learned C-R associations to novel stimuli from the same category is robust to changes in the response modality (Aim 1) and can be found in the absence of an overt response during training (Aim 2), even when the stimuli are pictures of everyday items and the classifications are common semantic categories.

Only the ANOVAs for the Enter and Speak modalities found that performance was improved when the test stimuli were repeated from the preceding learning block (Enter: mean RT = 1033 ms, errors = 3.9%; Speak: mean RT = 1100 ms, errors = 5.8%) relative to when they were novel (Enter: mean RT = 1107 ms, errors = 5.9%; Speak: mean RT = 1149 ms, errors = 7.5%). This was confirmed by a reliable main effect of Stimulus for both dependent variables in both learning response modalities (Enter RTs: *p*<.001, *BF*<0.001 ± 2.8%; Enter errors: *p*=.003, *BF*=0.068 ± 6.2%; Speak RTs: *p*<.001, *BF*=0.007 ± 7.5%; Speak errors: *p*=.010, *BF*=0.256 ± 5.2%). Conversely, the ANOVAs for the Listen condition found that test RTs were slightly faster, but participants made slightly *more* errors, when the stimuli were repeated from the preceding learning block (mean RT = 1136 ms, errors = 7.7%) relative to when they were novel (mean RT = 1146 ms, errors = 7.6%), but the main effect of Stimulus did not reach significance in either dependent variable (RTs: p=.297, *BF*=4.428 ± 2.8%; errors: p=.888, *BF*=6.996 ± 6.1%).

Taken together, these results indicate that S-R associations formed by speaking the response code readily transferred to the manual response modality (Aim 1), but the stimulus repetition benefit was limited to conditions where an overt response was required during the learning block (though that response need not be in the same modality as the response required at test). At least in the current paradigm, S-R associations formed in the absence of an overt response were not stable enough to affect subsequent performance where a manual response was required (note that the Bayesian analyses allowed us to reject the null hypothesis with moderate confidence). That is, as in Experiment 1, we were not able to replicate the key findings of Pfeuffer and colleagues (2017, 2018) in a paradigm that used a more complex task (Aim 3).

The Stimulus by Category interaction did not reach significance for any response modality in either dependent variable (ps>.07, BFs>1.6) with the exception of the RTs in the Listen modality (p=.007, BF=0.227 ± 2.9%). As in Experiment 1 (and in Longman, Milton, et al., 2018), when novel categories were introduced at test, test RTs were slower for those stimuli that had previously been classified under a different category (SsDcSr RT = 1172 ms) relative to altogether novel stimuli (DsDcSr RT = 1155 ms), suggesting some mild crosstalk. By contrast, when the categories were repeated between learning and test blocks, test RTs were faster for familiar stimuli (SsScSr RT = 1099 ms) relative to novel stimuli from the same category (DsScSr RT = 1137 ms), suggesting a mild stimulus repetition benefit.

General Discussion

The current experiments were designed to investigate the conditions under which the learning and spontaneous transfer of different kinds of associations between contexts is likely to occur. In both experiments we manipulated stimulus and category repetitions between the learning and test phases orthogonally to each other in order to directly compare the transfer of learned S-R and C-R associations to novel classifications/stimuli. We also manipulated the modality of the required response during the learning phase (a manual response was always required at test) to determine whether such associations learned in one response modality can transfer to a different modality (the Speak condition) and whether an overt response is even required at all (the Listen condition; the Enter condition was our baseline).

The results from both experiments indicate that old and novel (previously unseen) stimuli from trained categories can trigger a complex manual response when the required response during the learning phase was in a different modality to that required at test (Aim 1). Furthermore, the results from the Listen condition even showed that no overt practice was required at all (Aim 2). In both experiments, we also found strong evidence that test performance was improved in conditions where the categories were repeated from the previous learning block, whether the stimuli were repeated or novel, replicating our previous findings (Longman, Milton et al., 2018). Importantly, this pattern of results was not modulated by the required response modality during the learning blocks (i.e., it was found in the Listen and Speak conditions as well as the Enter condition). Combined, these findings suggest that C-R associations can be formed in the absence of procedural, practice-based memory traces. Verbal coding of responses, either acquired through speaking the codes or even through listening to them, is sufficient to learn associations that are later retrieved with a degree of automaticity. More generally, these results go some way toward softening the strong claims made by theories of learning/automaticity such as Instance Theory that proposes a high degree of specificity in learning (though, as noted in the Introduction, according to Instance Theory, the instance could indeed contain information about associative links beyond the specific stimulus/action). Note that the focus of the current experiments was on the early stages of learning (i.e., rule-based/algorithmic performance). The extent to which this pattern of results also applies to skilled (i.e., memory-based/automatic) performance remains to be seen, though some theorists would argue that repetition effects in the early stages of learning might be a first step toward automatization (e.g., Chein & Schneider, 2012; Logan, 1990).

Our findings provide strong support for C-R learning independent of the stimulus. Support for S-R learning independent of the category was more mixed. The results from Experiment 1 indicated that, when both the stimuli and responses were complex, evidence of a stimulus repetition benefit was largely limited to conditions where the categories were also repeated between the learning and test blocks. This pattern of results was observed in all modality conditions (Enter, Speak, Listen). That we found little evidence of S-R learning (and transfer) in any of the learning modality conditions would appear to contradict the results reported by Pfeuffer and colleagues (2017, 2018) who found that stable S-R associations could be formed in the absence of an overt response (Aim 3). However, the results of Experiment 1 are consistent with those reported by Longman, Milton, and colleagues (2018) who also failed to find S-R learning benefits in similar perceptual-categorization experiments. Possibly, the nature of the stimuli and categories played a role. Indeed, the results from the current Experiment 2, which used easily identifiable picture stimuli classified according to familiar semantic categories, were quite different. Although there was still strong evidence that learned C-R associations transferred to novel stimuli, there was also evidence that learned S-R associations transferred across classifications.

Critically, transfer of C-R associations was largely unaffected by changes in the response modality whereas transfer of S-R associations between classifications was limited to conditions where an overt response was made during training (though that response need not be in the test modality). Taken together, these results suggest that Pfeuffer and colleagues' observation that stable S-R associations can be formed in the absence of an overt response might be limited to very simple tasks/responses. However, participants in the current study were able to form C-R associations in the absence of an overt response in both experiments. The latter observation is consistent with Pfeuffer et al.'s claim that associative learning does not necessarily require overt practice, even if the tasks/responses are complex. We can speculate that this apparent contradiction might be due to the emphasis on the C-R bindings inherent in the instructions (cf. Longman, Milton et al., 2018), but more solid conclusions regarding this question will require further research.

The use of easily distinguishable pictures of everyday items might have made it easier to identify the test stimuli that were repeated from the preceding learning block when the classifications changed, as well as possibly making it easier to identify the novel stimuli introduced at test when the classifications were repeated from the preceding learning block. Thus, in Experiment 2, the use of easily distinguishable stimuli might have resulted in a greater emphasis on the exemplars (S-R bindings), whereas in Experiment 1, the use of highly artificial stimuli that were difficult to distinguish might have resulted in a greater emphasis on the rules (C-R bindings). This difference would likely be particularly prevalent in the early stages of learning where it is important to formulate a suitable strategy that will maximize the efficiency of learning. In other words, the easily identifiable (and familiar) stimuli used in Experiment 2 expedited the transition from rule-based/algorithmic learning to more memory-based/automatic processing by making the individual exemplar stimuli easy to memorize. Conversely, the artificial stimuli used in Experiment 1 presumably made memorizing the rules much easier than memorizing the individual exemplar stimuli. A conclusion that would seem to sit well with Logan's (1988) Instance Theory. Nonetheless, this shift in emphasis was apparently not enough to allow the formation of stable S-R associations in the absence of an overt response (cf. Pfeuffer et al., 2017, 2018; Aim 3).

Learned C-R associations (and S-R associations in Experiment 2) were robust to changes in the response modality, even when the responses were more complex than simple left/right key presses. The latter observation begs the question of how the (abstract) response in our study is represented – as a sequence of digits or a sequence of actions. Previous studies have already examined how responses might be represented in S-R associations (e.g., Dennis & Perfect, 2013; Horner & Henson, 2011; Liefooghe et al., 2012). However, these studies used simple responses (e.g., left vs. right), which makes it difficult to distinguish between semantic ("left") and motoric (left hand key press) coding. That performance was best in the Enter condition in the current experiments might indicate that there is a motor component to learning. However, the interactions with Modality failed to reach significance (and the Bayesian analyses found substantial evidence for the null hypothesis) in Experiment 1, and learning was also observed in the Speak and Listen conditions. Combined, these findings suggest that the stimuli/categories had become associated with more abstract (semantic) codes. However, the nature of the practice phase makes it difficult to determine whether participants coded the response as a sequence of digits (e.g., 2, 3) or (abstract[9]) motor actions (e.g., down, right) from the outset. If the participants coded the response as a sequence of digits then it should readily transfer to a different array of number keys (e.g., the keys along the top of a standard keyboard). On the other hand, if it was coded as a sequence of actions then it should readily transfer to a novel array of keys requiring the same sequence of actions (e.g., down, right). Further research is necessary to clarify this complex issue.

We attributed transfer effects in the Speak and Listen conditions to the retrieval of non-specific C-R (and S-R) associations. An alternative hypothesis is that improved performance was due to familiarity with the stimuli and categories. However, our prior work is inconsistent with this idea. In their Experiment 4, Longman, Milton and colleagues (2018) manipulated response repetitions (as well as stimulus and classification repetitions) between training and transfer blocks. Although they did not report the relevant inferential statistics, this manipulation allowed them to directly compare conditions where only the stimulus or only the category template was repeated from training to conditions where the S-R/C-R bindings were repeated from training – i.e., it allowed them to determine whether stimulus/category familiarity or S-R/C-R transfer was driving the effect of interest. Transfer performance was numerically better when the C-R bindings were repeated between training and transfer relative to when only the category templates were repeated (but not the stimuli or responses), suggesting that the effect was due to transfer of the learned C-R associations rather than familiarity with the category templates exclusively. Similarly, transfer performance was numerically better when the S-R bindings were repeated from training relative to when only the stimuli were repeated (but not the categories or the responses), suggesting that S-R transfer improved performance above that which could be explained exclusively by stimulus

[9] Note that participants could not perform a manual response during the Speak and Listen learning blocks because they had to clench a fist.

familiarity. In the current experiments, we did not manipulate response repetitions because it was not crucial to our main research questions and would have unnecessarily complicated an already complex design. Although it is possible to argue that the current experiments fail to control for this potential confound[10], our previous findings suggest that familiarity will only play a minor role in the paradigms used here.

Conclusion

To conclude, we have extended the findings of Longman, Milton et al. (2018) by demonstrating that S-R and C-R associations formed in one response modality readily transfer to a different response modality, even if no (overt) response is required at all during training. Additionally, we also extended their findings by demonstrating that transfer of C-R associations to novel stimuli is not limited to complex (perceptual) categories but can also be found when classifying pictures of everyday items according to semantic categories (though we also found good transfer of S-R associations across classification under these conditions). Furthermore, we have found a limit to the findings of Pfeuffer and colleagues (2017, 2018) by failing to replicate their central finding (S-R learning in the absence of an overt response) in our more complex tasks. Finally, we were able to demonstrate that highly abstract stimulus/category representations can become associated with highly abstract response representations even in the absence of an overt response during training. That these associations resulted in the triggering of a novel (previously unpracticed) response on detection of a novel (previously unseen) stimulus indicates the need to modify the traditional view that skilled performance develops via overt practice of a specific action in conjunction with a particular stimulus.

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

Funding: This work was supported by the European Research Council (grant number 312445 awarded to Frederick Verbruggen) and by a grant of the Deutsche Forschungsgemeinschaft (KI1388/5-1, Andrea Kiesel).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent: all persons gave their informed consent prior to their inclusion in the study.

Data Repository: All raw data files, R scripts (for data analysis), Matlab scripts (for data collection) and stimuli from both experiments are stored on the Open Science Framework data repository (https://osf.io/uw2bm/).

References

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

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

Cohen-Kdoshay, O., & Meiran, N. (2007). The representation of instructions in working memory leads to autonomous response activation: Evidence from the first trials in the flanker paradigm. *Quarterly Journal of Experimental Psychology,* 60 (8), 1140-1154.

Cohen-Kdoshay, O., & Meiran, N. (2009). The representation of instructions operates like a prepared reflex. *Experimental Psychology,* 56 (2), 128-133.

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.

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

Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology,* 79 (4), 481-492.

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.

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.

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.

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.

Kiesel, A., Wendt, M., & Peters, A. (2007). Task Switching: On the origins of response congruency effects. *Psychological Research, 71*, 117-125.

Kramer, A. F., Strayer, D. L., & Buckley, J. (1990). Development and transfer of automatic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 16 (3), 505-522.

Liefooghe, B., De Houwer, J., & Wenke, D. (2013). Instruction-based response activation depends on task preparation. *Psychonomic Bulletin and Review,* 20, 481-487.

Liefooghe, B., Wenke, D., & De Houwer, J. (2012). *Journal of Experimental Psychology:*

Learning, Memory and Cognition, 38 (5), 1325-1335.

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

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, 22, 1-35.

Longman, C. S., Liefooghe, B., & Verbruggen, F. (2018). How does the (re)presentation of instructions influence their implementation? Manuscript in preparation.

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.

Meiran, N., Cole, W. C., & Braver, T. S. (2012). When planning results in loss of control: Intention-based reflexivity and working memory. *Frontiers in Human Neuroscience*, 6 (104), 1-12.

Meiran, N., Pereg, M., Givon, E., Danieli, G., & Shahar, N. (2016). The role of working memory in rapid instructed task learning and intention-based reflexivity: An individual differences examination. *Neuropsychologia,* 90, 180-189.

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015a). The power of instructions: Proactive configuration of stimulus-response translation. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 41 (3), 768-786.

Meiran, N., Pereg, M., Kessler, Y., Cole, M. W., & Braver, T. S. (2015b). Reflexive activation of newly instructed stimulus-response rules: Evidence from lateralized readiness potentials in no-go trials. *Cognitive, Affective and Behavioral Neuroscience, I15 (2), 365-373.*

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *BayesFactor: Computation of Bayes factors for common designs (Version 0.9.11-1).*

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.

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.

Moutsopoulou, K., Yang, Q., Desantis, A., & Waszak, F. (2015). Stimulus-classification and stimulus-action associations: Effects of repetition learning and durability. *Quarterly Journal of Experimental Psychology,* 68 (9), 1744-1757.

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.

Pfeuffer, C.U., Hosp, T., Kimmig, E., Moutsopoulou, K., Waszak, F., & Kiesel, A. (2018). Defining stimulus representation in stimulus-response associations formed on the basis of task execution and verbal codes. *Psychological Research*, 82, 744–758.

Pfeuffer, C. U., Moutsopoulou, K., Pfister, R., Waszak, F., & Kiesel, A. (2017). *Journal of Experimental Psychology: Human Perception and Performance,* 43 (2), 328-347.

Pfeuffer, C. U., Moutsopoulou, K., Waszak, F., & Kiesel, A. (2018). Multiple priming instances increase the impact of practice-based but not verbal code-based stimulus-response associations. *Acta Psychologica, 184,* 100-109.

R Development Core Team, (2018). *R: A language and environment for statistical computing.* Vienna, Austria.

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.

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

Strube, M. (2006). SNOOP: A program for demonstrating the consequences of premature and repeated null hypothesis testing. *Behavior Research Methods*, 38 (1), 24-27.

Schuster, C., Hilfiker, R., Amft, O., Scheidhauer, A., Andrews, B., Butler, J., Kischka, U., & Ettlin, T. (2011). Best practice for motor imagery: A systematic literature review on motor imagery training elements in five different disciplines. *BMC Medicine,* 9, 75.

Verbruggen, F., McLaren, R., Pereg, M., & Meiran, N. (in press). Structure and implementation of novel task rules: A cross-sectional developmental study. *Psychological Science*.

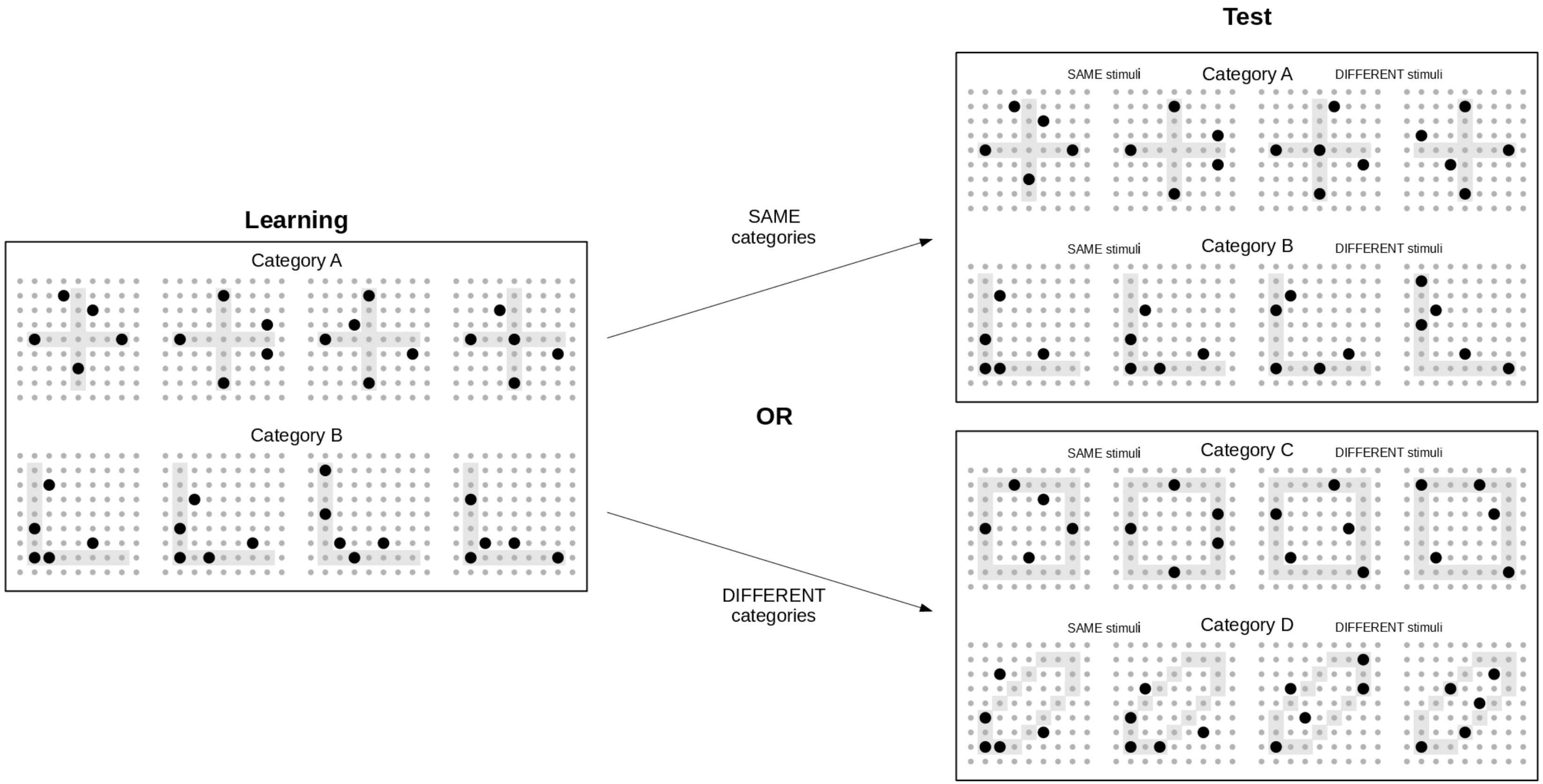


Figure 1
Overview of the experimental design and some example stimuli from each condition. Note that the category templates (presented here in grey) are to demonstrate how each exemplar stimulus belongs to the given category and were not displayed during the experimental trials. Novel stimuli and categories were introduced for each learning block. At test, the categories could either be novel or repeated from the preceding learning block. Whether the test categories were novel or repeated, half of the test stimuli were novel, and half were repeated from the preceding learning block.

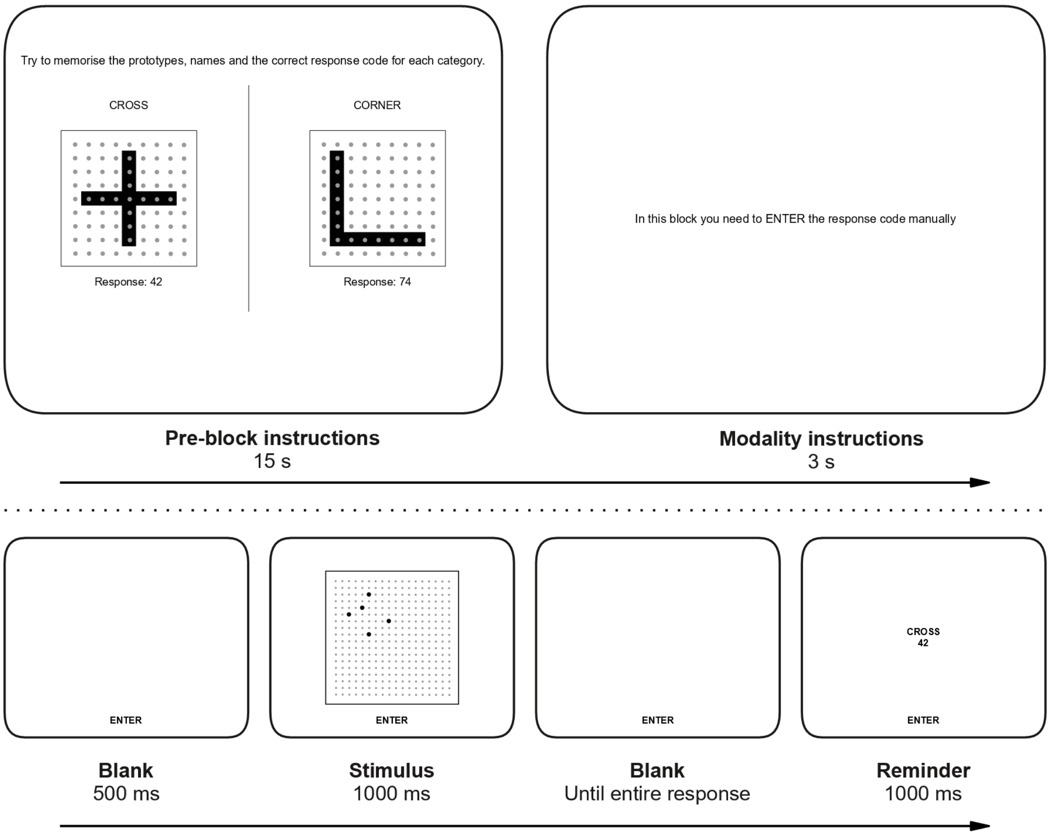


Figure 2
Pre-block instructions (top) and the timeline of a single trial (bottom) from Experiment 1. The procedure for Experiment 2 was identical except the instructions were displayed in German and the target stimuli were black drawings of everyday items classified according to semantic categories (see text for details).

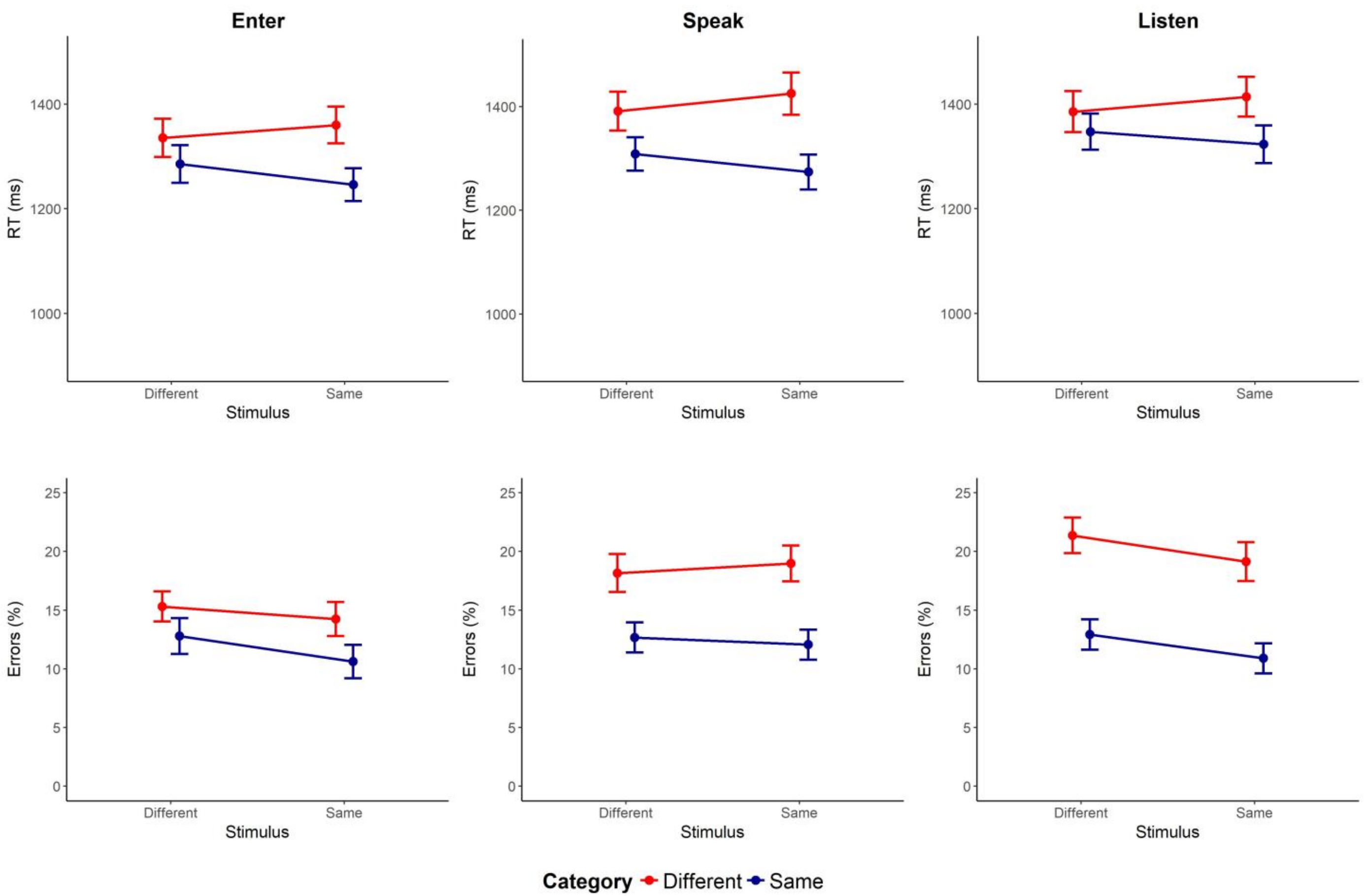


Figure 3
Mean RTs (top) and errors (bottom) from the test blocks for Experiment 1 plotted as a function of Stimulus and Category repetitions, separately for each learning response modality (see text for definitions). Error bars show the standard error of the mean.

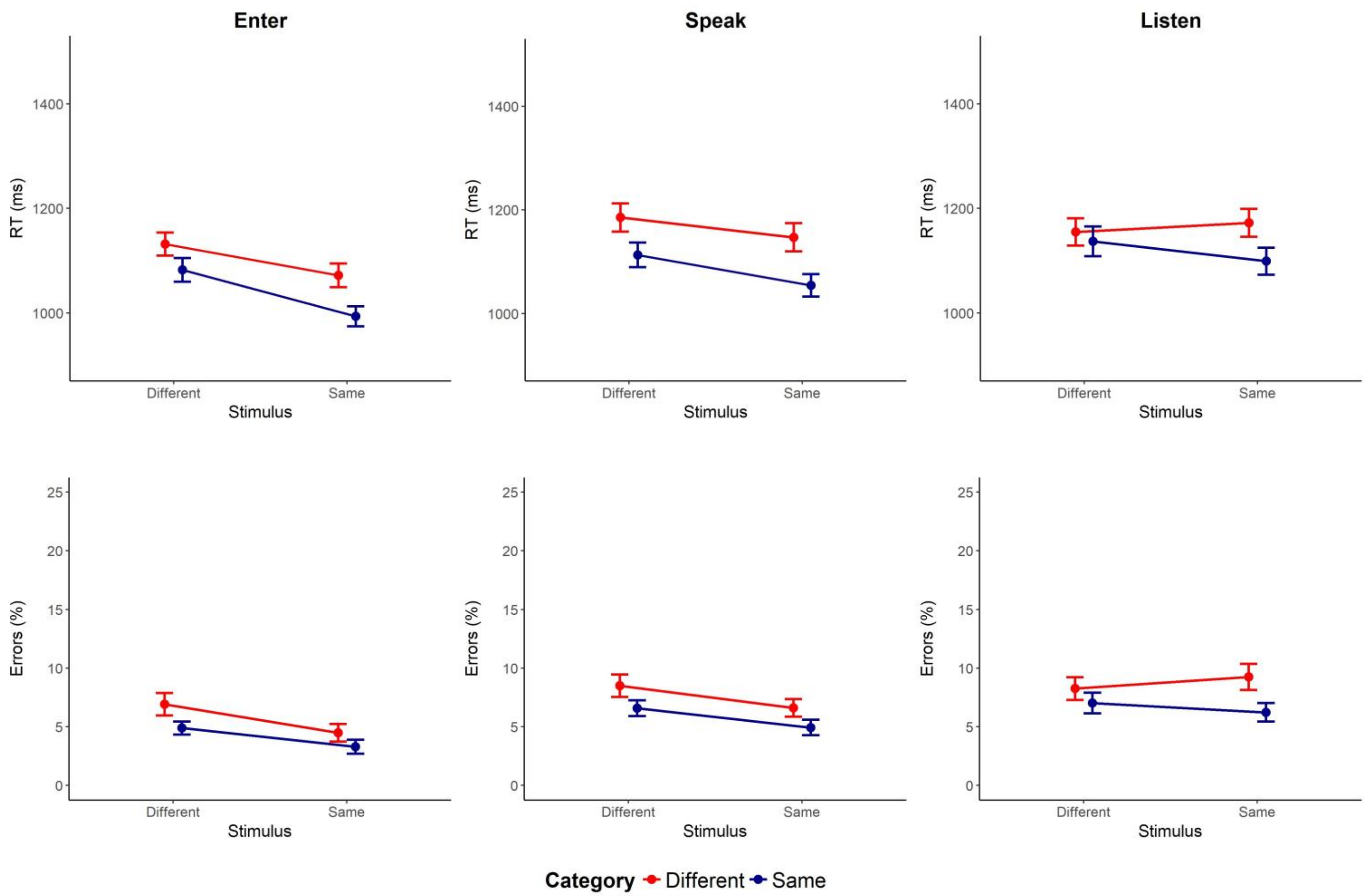


Figure 4
Mean RTs (top) and errors (bottom) from the test blocks for Experiment 2 plotted as a function of Stimulus and Category repetitions, separately for each learning response modality (see text for definitions). Error bars show the standard error of the mean.

Table 1: *Omnibus ANOVA Results for Experiment 1. Equivalent Bayes Factors are also Reported.*

	RT						Errors					
Effect	***DF***	***MSE***	***F***	***p***	***η2***	***BF***	***DF***	***MSE***	***F***	***p***	***η2***	***BF***
Modality	(2, 118)	63655.57	3.68	.028	0.026	0.019 ± 5.1%	(2, 118)	129.20	4.13	.018	0.017	0.259 ± 8.0%
Stimulus	(1, 59)	11165.49	0.05	.818	<0.001	12.161 ± 4.4%	(1, 59)	65.11	4.09	.048	0.004	3.492 ± 13.8%
Category	(1, 59)	44757.65	30.99	<.001	0.072	<0.001 ± 3.9%	(1, 59)	125.32	49.52	<.001	0.092	<0.001 ± 8.3%
Modality × Stimulus	(2, 118)	7600.20	0.21	.814	<0.001	36.571 ± 12.8%	(2, 118)	58.43	1.40	.250	0.003	14.898 ± 7.7%
Modality × Category	(2, 118)	38551.78	1.11	.334	0.005	8.807 ± 5.5%	(2, 118)	126.33	3.34	.039	0.014	0.503 ± 8.6%
Stimulus × Category	(1, 59)	11917.24	14.38	<.001	0.010	0.451 ± 5.6%	(1, 59)	71.40	0.38	.541	<0.001	9.035 ± 16.0%
Modality × Stimulus × Category	(2, 118)	6875.92	0.15	.859	<0.001	19.368 ± 5.3%	(2, 118)	74.03	0.15	.860	<0.001	15.372 ± 7.5%

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

Table 2: *Omnibus ANOVA Results for Experiment 2. Equivalent Bayes Factors are also Reported.*

	RT						Errors					
Effect	***DF***	***MSE***	***F***	***p***	***η2***	***BF***	***DF***	***MSE***	***F***	***p***	***η2***	***BF***
Modality	(2, 118)	20855.52	15.95	<.001	0.098	<0.001 ± 12.3%	(2, 118)	58.28	8.17	<.001	0.039	<0.001 ± 6.2%
Stimulus	(1, 59)	4646.15	76.06	<.001	0.054	<0.001 ± 11.9%	(1, 59)	25.74	10.55	.002	0.011	0.284 ± 7.3%
Category	(1, 59)	9401.83	78.47	<.001	0.107	<0.001 ± 11.9%	(1, 59)	55.01	11.16	.001	0.025	0.003 ± 6.7%
Modality × Stimulus	(2, 118)	4922.16	12.67	<.001	0.020	0.065 ± 12.1%	(2, 118)	27.71	2.90	.059	0.007	3.947 ± 6.0%
Modality × Category	(2, 118)	11126.36	1.80	.169	0.006	4.578 ± 14.0%	(2, 118)	37.15	0.11	.896	<0.001	33.681 ± 15.4%
Stimulus × Category	(1, 59)	5909.71	9.35	.003	0.009	0.521 ± 14.5%	(1, 59)	31.46	0.08	.777	<0.001	8.854 ± 11.4%
Modality × Stimulus × Category	(2, 118)	5185.48	0.93	.396	0.002	10.773 ± 14.5%	(2, 118)	22.12	1.28	.281	0.002	8.161 ± 5.5%

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

Table 3: *ANOVA Results by Training Response Modality for Experiment 2. Equivalent Bayes Factors are also Reported.*

	Enter											
	RT						**Errors**					
Effect	***DF***	***MSE***	***F***	***p***	***η2***	***BF***	***DF***	***MSE***	***F***	***p***	***η2***	***BF***
Stimulus	(1, 59)	4648.70	71.19	<.001	0.281	<0.001 ± 2.8%	(1, 59)	25.04	9.75	.003	0.052	0.068 ± 6.2%
Category	(1, 59)	5797.75	42.27	<.001	0.224	<0.001 ± 3.6%	(1, 59)	33.08	4.71	.034	0.034	0.350 ± 5.6%
Stimulus*Category	(1, 59)	3932.68	3.27	.076	0.015	1.651 ± 2.1%	(1, 59)	17.57	0.59	.444	0.002	4.076 ± 5.5%

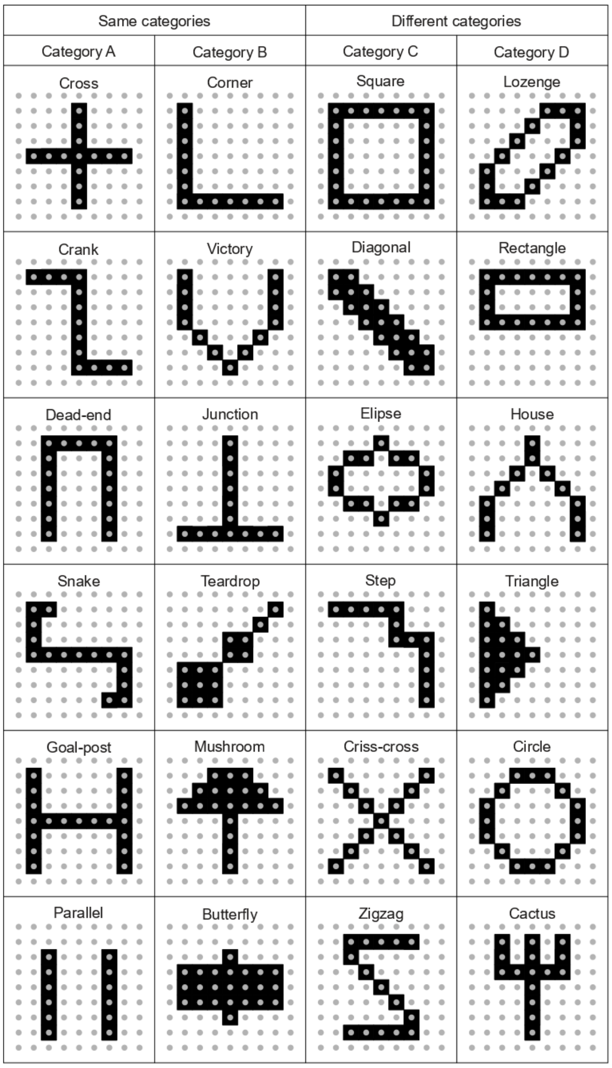
	Speak											
	RT						**Errors**					
Effect	***DF***	***MSE***	***F***	***p***	***η2***	***BF***	***DF***	***MSE***	***F***	***p***	***η2***	***BF***
Stimulus	(1, 59)	4196.97	33.58	<.001	0.078	0.007 ± 7.5%	(1, 59)	26.36	7.11	.010	0.038	0.256 ± 5.2%
Category	(1, 59)	17439.33	23.34	<.001	0.196	<0.001 ± 7.2%	(1, 59)	29.58	6.58	.013	0.039	0.224 ± 5.1%
Stimulus*Category	(1, 59)	6593.43	0.97	.328	0.004	3.513 ± 7.6%	(1, 59)	24.55	0.04	.851	<0.001	4.982 ± 4.6%

	Listen											
	RT						**Errors**					
Effect	***DF***	***MSE***	***F***	***p***	***η2***	***BF***	***DF***	***MSE***	***F***	***p***	***η2***	***BF***
Stimulus	(1, 59)	5644.80	1.11	.297	0.005	4.428 ± 2.8%	(1, 59)	29.77	0.02	.888	<0.001	6.996 ± 6.1%
Category	(1, 59)	8417.48	14.95	<.001	0.097	0.001 ± 5.0%	(1, 59)	66.65	4.08	.048	0.034	0.329 ± 6.1%
Stimulus*Category	(1, 59)	5754.56	7.93	.007	0.038	0.227 ± 2.9%	(1, 59)	33.60	1.43	.237	0.006	2.971 ± 6.0%

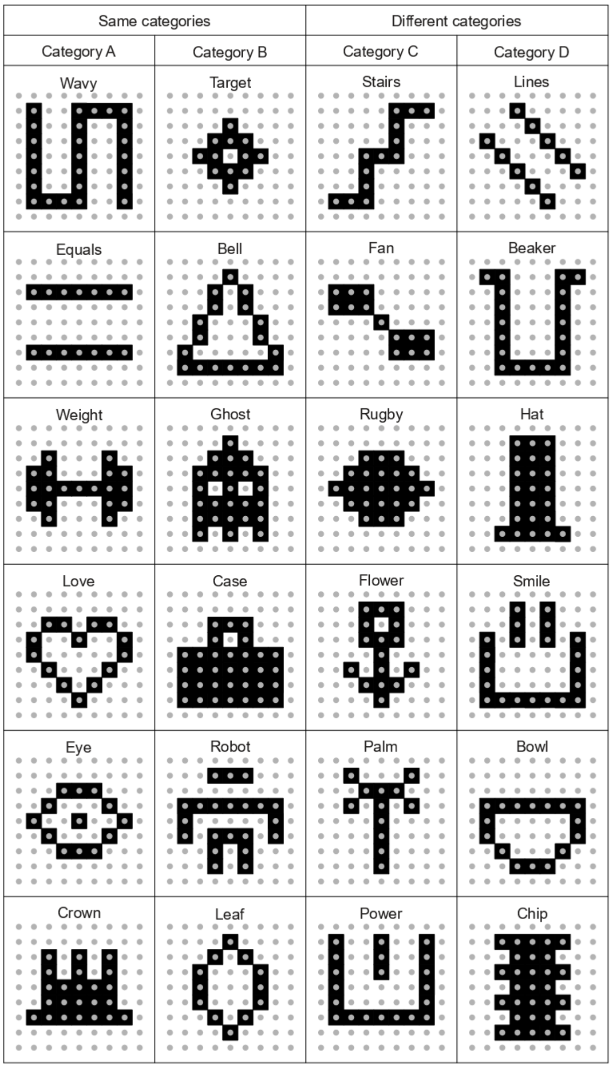
Note: Bayes factors indicate whether removal of the effect/interaction from the model would materially impair its fit. Thus, Bayes factors<1 indicate that the effect/interaction is an important contributor to the model.

Appendix 1

Templates for all categories used in Experiment 1. The category labels were used to provide regular reminders to participants. Each row in the table shows the category labels and templates for the categories that were used during a given learning-test block pair. Labels/templates in the columns marked 'Same categories' (Categories A and B) were used during the learning block as well as during the subsequent test block when the categories were repeated from the preceding learning block. Labels/templates in the columns marked 'Different categories' (Categories C and D) were only used during the test block for conditions where novel categories were introduced at test.



Same categories		Different categories	
Category A	Category B	Category C	Category D
Cross	Corner	Square	Lozenge
Crank	Victory	Diagonal	Rectangle
Dead-end	Junction	Elipse	House
Snake	Teardrop	Step	Triangle
Goal-post	Mushroom	Criss-cross	Circle
Parallel	Butterfly	Zigzag	Cactus



Same categories		Different categories	
Category A	Category B	Category C	Category D
Wavy	Target	Stairs	Lines
Equals	Bell	Fan	Beaker
Weight	Ghost	Rugby	Hat
Love	Case	Flower	Smile
Eye	Robot	Palm	Bowl
Crown	Leaf	Power	Chip

Appendix 2

Category labels and descriptions for all categories used in Experiment 2. As in Experiment 1, the category labels were used to provide regular reminders to participants. Each row in the table shows the category labels and descriptions for the categories that were used during a given learning-test block pair. Labels/descriptions in the columns marked ‘Same categories’ (Categories A and B) were used during the learning block as well as during the subsequent test block when the categories were repeated from the preceding learning block. Labels/descriptions in the columns marked ‘Different categories’ (Categories C and D) were only used during the test block for conditions where novel categories were introduced at test.

Same categories		**Different categories**	
Category A	**Category B**	**Category C**	**Category D**
SWIMS Usually swims in water	WALKS Usually walks on land	SMALL Smaller than a shoe box	BIG Bigger than a shoe box
STATIONERY An item of stationery	TOOLS A hand tool	MECHANICAL Has moving parts	NON-MECHANICAL Has no moving parts
CLOTHES An item of clothing	ACCESSORIES A fashion accessory	METAL Made from metal	TEXTILES Made from textiles
LIVING A living organism	MAN-MADE A man-made item	WHEELS The item has wheels	WINGS The item has wings
CUTLERY An item of cutlery	CROCKERY An item of crockery	BLUNT The item is blunt	SHARP The item is sharp
VEGETABLE An edible vegetable	FRUIT An edible fruit	SOFT Food that is soft	CRUNCHY Food that is crunchy
MUSIC Musical instrument	SPORT Sports equipment	PLASTIC Made from plastic	WOOD Made from wood
NEW The item is new	OLD The item is old	SOUND A source of sound	LIGHT A source of light
SCIENCE Science equipment	ART Art equipment	OPAQUE The item is not transparent	TRANSPARENT The item is transparent
TOY A toy to play with	GAME A game to play	STRAIGHT The item has straight edges	ROUND The item has a round edge
FLOATS The item floats in water	SINKS The item sinks in water	KITCHEN Commonly found in the kitchen	BATHROOM Commonly found in the bathroom
SOLID The item is a solid	LIQUID The item is a liquid	HOT The item is hot	COLD The item is cold