Citation for final published version:

O'Donoghue, Ellen M. and Wasserman, Edward A. 2025. A comparative investigation of task-switching performance in category learning paradigms. Journal of Experimental Psychology: Learning, Memory, and Cognition 10.1037/xlm0001564

Publishers page: https://doi.org/10.1037/xlm0001564

A Comparative Investigation of Task Switching Performance in Category Learning Paradigms

Ellen M. O’Donoghue and Edward A. Wasserman

The University of Iowa

Ellen M. O’Donoghue and Edward A. Wasserman, Department of Psychological and Brain Sciences, The University of Iowa.

Correspondence concerning this article should be addressed to Ellen O’Donoghue, CUBRIC, Maindy Rd., Cardiff University, Cardiff, UK. CF24 4HQ. E-mail: odonoghuee1@cardiff.ac.uk

Abstract

Recent research has demonstrated that task switching – which has traditionally been attributed to declarative rule use and executive control – can also be supported by associative learning mechanisms. However, whereas declarative task switching is characterized by large switch costs and small congruency effects, associative mechanisms may produce small (or no) switch costs and large congruency effects. Here, we asked whether humans (who possess both declarative and associative learning mechanisms) and pigeons (which have thus far shown no evidence of possessing declarative learning mechanisms) would display different patterns of performance when switching either between two rule-based (RB) subtasks, which should encourage declarative rule use, or between two information-integration (II) subtasks, which should encourage associative learning. The pigeons showed no switch costs in either task condition, consistent with the view that they depend entirely on associative mechanisms to solve both RB and II tasks. Conversely, the humans showed strong switch costs in both task conditions. These data raise two possibilities: (1) that human learners may have used declarative mechanisms to solve both RB and II tasks, and (2) that among humans, associative learning mechanisms might not reliably preclude switch costs. The theoretical implications of each possibility are discussed.

Keywords: Task-switching, categorization, pigeons, RB-II, dual-systems

Several prominent theories of learning and categorization have argued that humans have two different learning mechanisms – one declarative, one associative – with the potential to promote adaptive responding (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; McLaren et al., 2019; see also Minda, Roark, Kalra & Cruz, 2024). In comparison to declarative processes, which may be restricted to select species (including humans and possibly some nonhuman primates), associative processes are evolutionarily ancient and conserved across species (Smith et al., 2012; see also Pontes et al., 2020). Humans' impressive cognitive achievements are often attributed to our capacity for declarative processing, and by contrast, some authors have drawn a sharp dividing line between associative learning and more "complex" forms of cognition (such as concept formation, language, and executive control; e.g., Allen & Bekoff, 1995; Marcus, Vijayan, Rao, & Vishton, 1999; Mulcahy & Call, 2006).

Nevertheless, several lines of inquiry stemming from machine learning (Silver et al., 2016; Silver et al., 2017) and cognitive psychology (Lind, 2018; McMurray, Horst, & Samuelson, 2012; Sloutsky, Yim, Yao, & Dennis, 2017; Wasserman, Kain, & O'Donoghue, 2023) have demonstrated that associative learning can support highly flexible, goal-directed behaviour across varied cognitive domains. In line with this perspective, researchers have found that task switching – which has traditionally been viewed with the ambit of executive control and declarative processing (e.g., Monsell, 2003) – can likewise be accomplished through associative mechanisms (Castro & Wasserman, 2016; Meier, Lea, Forrest, Angerer, & McLaren, 2013; Meier, Lea, & McLaren, 2016; Forrest, Elchlepp, Monsell, & McLaren, 2012; Li, Li, Liu, Lages, & Stoet, 2019a; O'Donoghue & Wasserman, 2021).

In a typical task switching paradigm, participants concurrently perform two or more tasks involving the same stimulus set. On every trial, a *task cue* is presented to signal which contingencies of reinforcement are in force on that trial. For example, a hypothetical participant might have to classify a number according to either quantity (high/low) or parity (odd/even), contingent on the color of a background cue. These two tasks would be intermixed within each session, thereby requiring the participant to shift their classification strategies on a trial-by-trial basis.

As in the above example, most task switching paradigms explore the effect of switching between declarative decision rules; indeed, participants are often directly informed of the rules prior to beginning the experiment (c.f., Meier et al., 2013). Under these circumstances, human participants evidence at least two characteristic patterns of responding.

First, participants display small *congruency effects*: they are slightly, but reliably faster (and sometimes more accurate) to respond to stimuli which require the same response regardless of the task cue, relative to stimuli which require different responses depending on the task cue (Kiesel, Wendt, & Peters, 2007). For example, if “odd numbers” and “numbers < 5” are assigned to one response key, whereas “even numbers” and “numbers > 5” are assigned to the second response key, then “3” is a congruent stimulus (because it requires the same response in both tasks) and “7” is incongruent (because it requires a different response in each task).

Second, participants display large *switch costs*: when trial *N* involves a different decision rule than trial *N-1*, response times increase and/or accuracies decrease relative to when the same decision rule repeats across trials (Monsell, 2003; Samavatyan & Leth-Steensen, 2009). These switch costs are often attributed to the need to discard the executive *task set* that was appropriate for trial *N-1*, and to recruit the task set that is appropriate for trial *N.* Switch costs have been framed as a hallmark of executive functioning (e.g., Monsell, 2003): they persist across a wide variety of paradigms and populations (Ardiale, Hodzik, & Lemaire, 2012; Ellefson, Shapiro, & Chater, 2006); they are not eliminated (although they can be reduced) through advanced preparation (Nieuwenhuis & Monsell, 2002), extended practice (Stoet & Snyder, 2007), or expertise (Slama, Rebillon, & Kolinsky, 2017); and they may have adaptive value because they promote sustained attention to ongoing tasks (Li, Li, Lages, & Stoet, 2017) while they shield participants from irrelevant sources of variance (Dreisbach & Haider, 2008).

Associative Task Switching

If switch costs are a byproduct of executive control, then they should be eliminated (or at least strongly reduced)[1] in paradigms that do not encourage declarative rule use. Although associative task switching is less well-studied than declarative task switching, studies involving both human associative learners (e.g., Forrest et al., 2012; Li et al., 2019; Meier et al., 2013) and nonhuman species which may lack human-like executive functioning (e.g., Castro & Wasserman, 2016; Meier et al., 2013; Meier et al., 2016; O'Donoghue, Broschard, & Wasserman, 2020; O'Donoghue & Wasserman, 2021) are thus far consistent with this prediction. In contrast to rule users, who typically produce large switch costs relative to small congruency effects, associative learners typically produce small (or no) switch costs relative to large congruency effects (e.g., Forrest et al., 2012; Li et al., 2019a; Meier et al., 2013; see Forrest et al., 2012, for proof-of-concept that a purely associative mechanism can produce small switch costs).

Comparative investigations of associative task switching are complicated by the fact that humans are strongly biased toward declarative rule use (Ashby et al., 1998; Smith et al., 2012), which makes the contributions of associative mechanisms extremely difficult to isolate (McLaren et al., 2019). In lieu of explaining task rules to human participants (as is standard in the declarative literature), most associative task switching paradigms either require research participants to memorize a comprehensive list of cue-stimulus-response (CSR) mappings (e.g. Li et al., 2019a) or to use small stimulus sets that are likewise amenable to exemplar memorization (Forrest et al., 2012; Meier et al., 2013). In some cases, extra precautions are taken to ensure that task rules are non-discoverable; for example, Li, Li, Liu, Lages, and Stoet (2019b) investigated task switching performance among English-speaking participants tasked with categorizing Chinese numerals.

---

[1] Note that switch costs can be incurred not just by switching *tasks,* but also by switching *cues* (e.g., Jost, De Baene, Koch, & Brass, 2013). We revisit this issue in the General Discussion.

Whether such memorization is truly associative remains a matter of debate. Some authors have asserted that CSR memorization primarily reflects the contributions of associative processes (Bower & Winzenz, 1970), whereas others have argued that associative learners should not have conscious access to their memory representations (e.g., Edmunds, Wills, & Milton, 2016; Smith et al., 2012), meaning that effortful memorization would constitute a declarative learning strategy.

With this debate in mind, we aimed to compare humans' declarative and associative task switching performance using a paradigm that *encourages* associative learning (by making task rules difficult to verbalize) while also *discouraging* CSR memorization (through the use of trial-unique stimuli). To do so, we compared task switching performance among participants trained to solve rule-based (RB) and information-integration (II) tasks like those depicted in Figure 1.

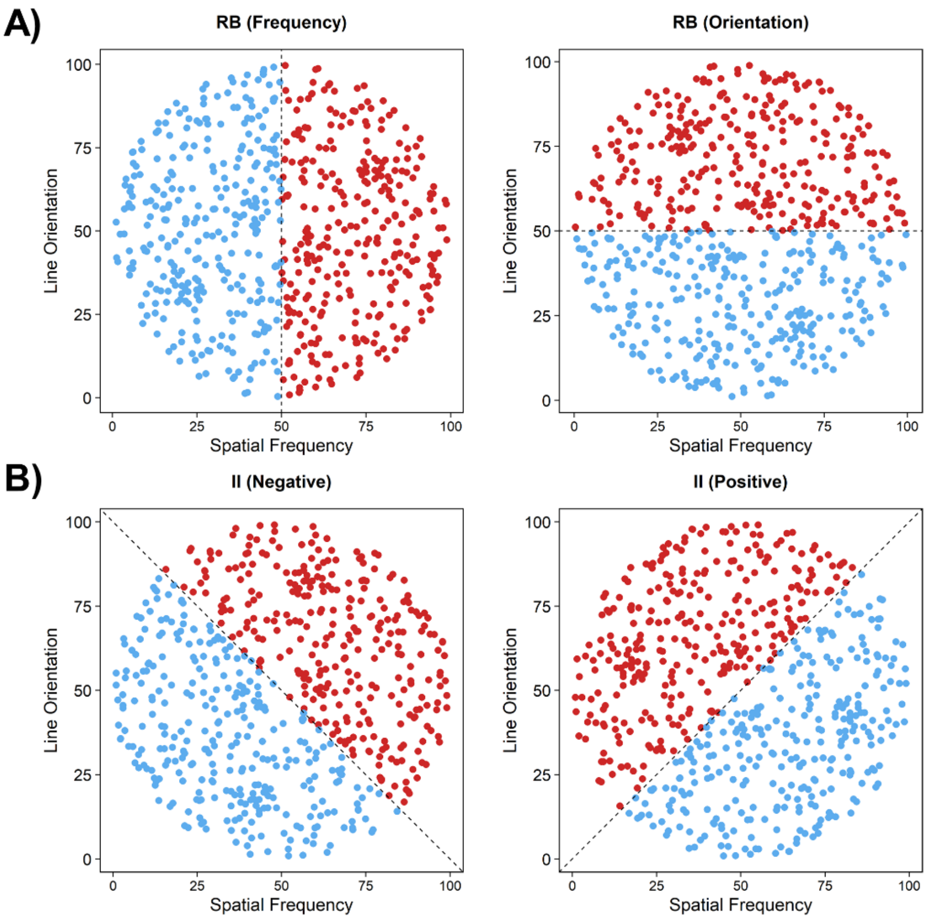

Figure 1. Sample rule-based (A) and information-integration (B) stimulus distributions. The color of the points denotes category assignment, and the dashed lines denote the optimal decision bounds. Note that, although the stimulus distributions shown here are circular and uniformly sampled, most RB-II research has instead used bivariate normal distributions (e.g., Crossley, Roeder, Hélie, & Ashby, 2018; Smith et al., 2012). In our laboratory, we prefer to use circular distributions because they ensure that both tasks and all four subtasks are sampled from shared regions of the stimulus space (O'Donoghue et al., 2020; O'Donoghue & Wasserman, 2021).

RB and II tasks are created by sampling two categories from the same bidimensional stimulus space, often comprising sinusoidal gratings that vary in spatial frequency and line orientation (e.g., Crossley et al., 2018; Smith et al., 2012). In our RB tasks (Figure 1A), the optimal decision bound that separates the two categories lies perpendicular to a single dimension, rendering only that dimension relevant for categorization. Because these decision bounds are readily verbalizable (e.g., "if the lines are angled at less than 45°, then the stimulus belongs to category 'A'; otherwise, it belongs to category 'B'"), the RB tasks encourage declarative rule use (Ashby & Valentin, 2017).

By contrast, the diagonal decision bounds that delineate the II distributions are not easy to verbalize. To solve the II tasks, participants must integrate information about *both* spatial frequency and line orientation, which is often assumed to encourage associative processing (e.g., Ashby & Valentin, 2017; Smith et al., 2012).

Supporting the assertion that RB and II tasks may engage different learning mechanisms, human participants reliably learn RB tasks more quickly than II tasks. This discrepancy is often taken to reflect the fact that II learners begin by testing declarative decision rules, but must then abandon those rules to achieve task mastery (see Ashby & Valentin, 2017, for a review of the RB advantage alongside related RB-II dissociations). Critically, at least two nonhuman species believed to possess the associative system, but to lack the declarative system – namely, pigeons (O'Donoghue, et al., 2020; Smith et al., 2011; Smith et al., 2012) and rats (Broschard, Kim, Love, Wasserman, & Freeman, 2019) – show no differences in RB and II learning speed, suggesting that the disparities observed in the human literature cannot be explained by task difficulty (see also Ashby & Valentin, 2017; c.f., Nosofsky, Stanton, & Zaki, 2005; Zaki & Kleinschmidt, 2014)[2].

To date, very few authors have examined humans' task switching performance in the context of RB and II categorization. Meier et al. (2013) trained participants to switch between simplified versions of the two RB subtasks depicted in Figure 1A. After trial-and-error learning, participants who successfully verbalized the rules governing each subtask displayed strong switch costs and small congruency effects, consistent with declarative rule use. By contrast, participants who did not verbalize either categorization rule displayed no switch costs and strong congruency effects, consistent with associative learning. However, Meier et al.'s (2013) stimulus set comprised just four unique stimuli, which might have encouraged exemplar memorization even though their participants were not explicitly instructed to memorize the stimuli.

Separately, Crossley et al. (2018) asked whether participants could simultaneously learn and switch between an RB subtask and an II subtask involving randomly sampled stimuli, under the assumption that this procedure necessitates trial-by-trial switching between declarative and associative mechanisms (see also Erickson, 2008). Such switches proved to be difficult, but possible. In addition, RB-II switches were more difficult than switches between a unidimensional RB subtask and a bidimensional RB task, suggesting that switching between declarative and associative mechanisms may be more costly than switching between two declarative rules.

On the other hand, we were particularly interested in conditions that were not examined in any of this prior work: namely, requiring participants to switch between two unidimensional, randomly-sampled RB subtasks (Figure 1A), which should each depend on declarative rule use, or

---

[2] Beyond this basic *RB advantage,* RB and II tasks have yielded numerous behavioral and neurobiological dissociations suggesting that they engage different learning mechanisms (for a review, see Ashby & Valentin, 2017). However, several of these dissociations have either failed to replicate (e.g., Edmunds, Wills, & Milton, 2019; Newell, Dunn, & Kalish, 2010) or may have been confounded by other factors (e.g., Stephens & Kalish, 2018; Newell, Moore, Wills, & Milton, 2013). A full review of the debate surrounding RB and II tasks lies beyond the scope of the present report; however, we revisit this issue in the *General Discussion*.

requiring participants to switch between two bidimensional, randomly-sampled II subtasks (Figure 1B), which should each encourage associative mechanisms. If human learners engaged the optimal associative strategy to solve the two II subtasks, then we expected that they would show significantly weaker switch costs relative to human rule-users trained to switch between the two RB subtasks.

To determine whether human learners arrived at the optimal strategies for each task condition, we considered the concordance between two manipulation checks: decision bound modelling (DBM; Ashby & Gott, 1988) and self-reported strategy use (e.g., Edmunds et al., 2016). Additionally, and in effort to better discern which elements of humans' performance might be mediated by associative mechanisms (that are shared across species), and which elements might be mediated by declarative mechanisms (that may be unique to select species, including humans and possibly some nonhuman primates; Smith et al., 2012), we conducted a comparative investigation involving both humans (Experiment 1) and pigeons (Experiment 2A, wherein we trained pigeons using the same general approach used in Experiment 1, and Experiment 2B, wherein we asked whether differences in stimulus processing might explain our observed interspecies differences).

Experiment 1

In Experiment 1, we trained human participants to switch either between the two RB subtasks (the RB-RB group) or between the two II subtasks (the II-II group) using the full stimulus distributions depicted in Figure 1. If most participants converged on the optimal learning mechanisms for each task condition (declarative rule use in the RB-RB group; associative learning in the II-II group), then we suspected that participants in the RB-RB group would show significantly stronger switch costs than participants in the II-II group.

Method

Participants. We recruited 40 participants (mean age = 43.46, SD = 12.31) from Amazon's Mechanical Turk, each of whom completed Experiment 1 on a personal computer in exchange for monetary compensation ($10 USD). Twenty participants were randomly assigned to each of the RB-RB and II-II task conditions. All participants provided informed consent, and all experimental procedures were approved by the Human Subjects Office at The University of Iowa.

As described in the *Results*, we subsequently excluded 14 participants who did not meet our learning criterion (5 participants excluded from the RB-RB group, and 9 participants excluded from the II-II group, yielding a final RB-RB group N = 15 and a final II-II group N = 11). The mean age across the remaining 26 participants was 43.10 (SD = 11.95).

A post hoc sensitivity analysis conducted in G*Power revealed that, with combined N = 26 and a two-tailed, six-predictor model[3], we had 80.00% power to detect regression coefficients for which $f^2 = 0.34$ and 90.00% power to detect regression coefficients for which $f^2 = 0.45$. Critically, the effect sizes obtained in past research exceed these thresholds: in Li et al. (2019b), the interaction between task condition (declarative vs. associative) and trial type (stay vs. switch) on reaction time-based switch costs had an $\eta^2_p = 0.43$ (converted $f^2_p = 0.43 / (1 - 0.43) = 0.75$), while in Forrest et al. (2012), the effect of task condition (declarative vs. associative) on reaction-time based switch costs had an $f^2 = 0.53$. Thus, our final sample size was sufficiently powered to detect differences in switch cost magnitude as a function of task condition (should such differences exist using the RB and II tasks deployed here).

Stimuli. We generated the stimuli and programmed the experiment using JavaScript and the jsPsych library (de Leeuw, 2015). In each of the RB-RB and II-II task conditions, the stimuli were sinusoidal gratings that varied in spatial frequency and line orientation. These stimuli were randomly and uniformly sampled from the normalized circular distributions depicted in Figure 1, with raw

---

[3] As detailed in the *Results*, our statistical models included three individual predictors (task condition, congruency, and trial type) as well as all possible interactions between them (two two-way interactions and one omnibus three-way interaction).

spatial frequency values ranging from 0.01 to 0.10 cycles per pixel and raw line orientation values ranging from 0-90°.

Phases 1 and 2: Individual Subtask Training. All participants were independently trained on two individual subtasks [the frequency (RBF) and orientation (RBO) subtasks in the RB-RB group, or the negative (IIN) and positive (IIP) subtasks in the II-II group; Figure 1] before progressing to the critical task switching phase. The order in which the subtasks were trained was counterbalanced across participants. Before beginning the first subtask, all participants were told that they would "learn one way of sorting the objects", and before beginning the second subtask, all participants were told that they would "learn a different way of sorting the objects". In both phases, learning progressed through trial-and-error; no further task instructions were provided. Participants were asked to respond as quickly and accurately as possible.

Each trial began with the presentation of a 250 x 250 px circular task cue in the center of the screen (Figure 2A). This task cue could be either blue (RGB value: [0, 0, 255]) or red (RGB value: [255, 0, 0]), with cue-subtask relationships randomized across participants. The participant then had to press the space bar, at which point the 200 x 200 px trial exemplar was overlaid on top of the task cue. Once the trial exemplar appeared, participants were immediately allowed to categorize it by pressing either the 'q' or 'p' key on their keyboard. Category-response assignments were counterbalanced across the full sample. Participants were provided with feedback 0.5 s after making their response; either the word "correct" or "error" appeared on-screen, and remained visible for 1.0 s. The next trial then began automatically after a variable 0.8-1.2 s delay. No correction trials were provided.

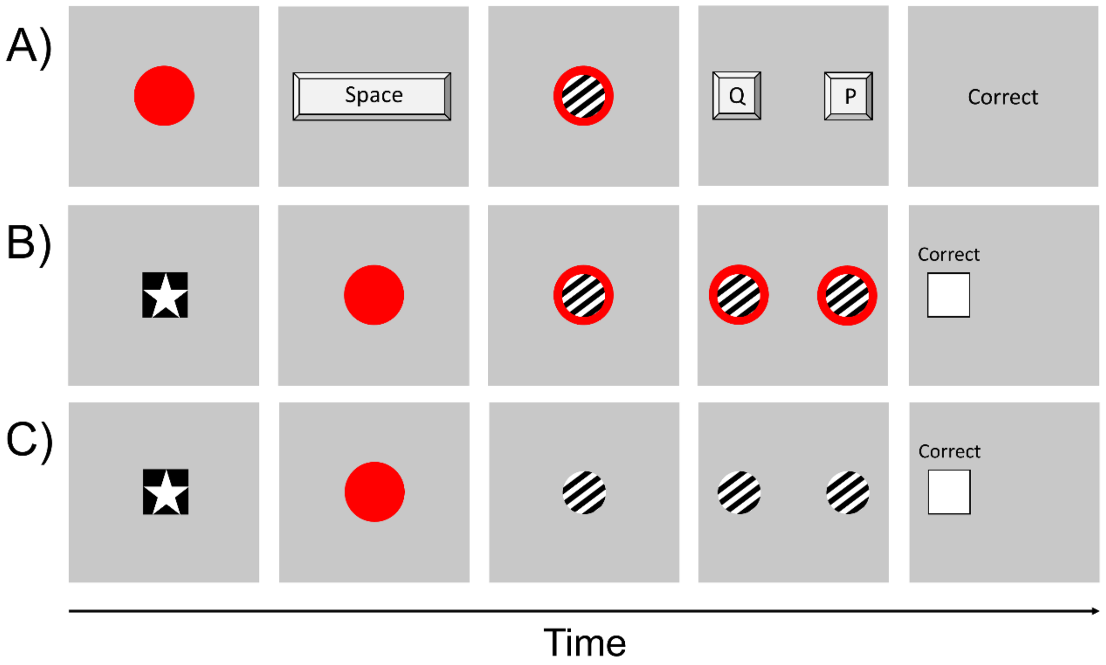

Figure 2. Trial structures used in Experiments 1 and 2. The top row (A) depicts the trial structure given to the humans in Experiment 1; the middle row (B) depicts the trial structure given to the pigeons in Experiment 2A; the bottom row (C) depicts the trial structure given to the pigeons in Experiment 2B. In panels B and C, the word "correct" is included for illustrative purposes only.

Training on each individual subtask continued either until the participant reached an 85% criterion (at least 85% overall accuracy in a 30-trial moving window) or to a maximum of 240 trials. For the purposes of our analyses, we classified participants as "learners" only if they met the 85% criterion separately on each individual subtask. We classified participants as "nonlearners" if they did not meet the 85% criterion on at least one subtask.

After finishing each individual subtask, participants were prompted to self-report how they had solved that subtask. ("In this phase, what do you think determined which response was correct? What approach(es) did you take to sorting the objects? Please be as specific as possible.")

Phase 3: Task switching. After receiving independent training on each of the two subtasks in the manner described above, participants progressed to the critical task switching phase. At the beginning of this phase, all participants were told that, having “learned two different ways of sorting the objects”, they would now have to “put them together”. Participants were reminded of the counterbalanced color-subtask mappings that they had learned in Phases 1 and 2. For example, a sample participant would be told: “when the background is blue, you should sort the objects as you did in Phase 1”, and “when the background is red, you should sort the objects as you did in Phase 2”.

The trial sequence was identical to that described in Phases 1 and 2. All participants completed 480 trials in the task switching phase. As did Meier et al. (2013), we partially randomized the trial order such that a subtask switch occurred on one third of all trials.

At the end of the task switching phase, participants were given two additional self-report prompts. The first asked whether their approaches to sorting the objects had changed in the task switching phase (“Did your approach to sorting the objects change in this final phase? If so, how?”), while the second asked whether they experienced any distractions during the study [“Were you distracted while you completed the task (e.g., by using your phone)?”]. No participants reported meaningful changes in their approach during the task switching phase nor did they report substantial distractions.

Results and Discussion

All data and scripts used for analysis are available via https://osf.io/3265r/ (O’Donoghue, 2023). This study was not preregistered.

Learning Speeds. To compare learning speeds in each of the RB-RB and II-II groups, we first excluded 14 nonlearners (five from the RB-RB group, and nine from the II-II group), defined as participants who did not meet the 85% criterion on at least one subtask during Phases 1 and 2 (see Method). After exclusion, there remained 15 participants in the RB-RB group, and 11 participants in the II-II group (total analyzable $N = 26$). We opted to exclude nonlearners from our analyses because, given that they did not reach criterion on at least one subtask, they presumably had not adopted the optimal strategy for that subtask (which would in turn muddle the group-level analyses).

Figure 3A depicts the average learning curves in each task condition (RB-RB vs. II-II), separated by subtask. For easier visualization, we divided each individual participant's data into 10 blocks (mean number of trials per block = 6.60; SD = 3.89). On average, participants acquired all four subtasks quite quickly – an impressive feat given the difficulty of our arranged stimulus distributions (Figure 1). As is typical of RB and II paradigms (see Ashby & Valentin, 2017), learning was numerically faster in each of the two RB subtasks (RBF: mean number of trials = 47.53; SD = 25.43; RBO: mean number of trials = 59.47; SD = 29.80) relative to each of the two II subtasks (IIN: mean number of trials = 81.91; SD = 40.84; IIP: mean number of trials = 84.27; SD = 47.66). An independent-samples *t*-test with overarching task condition (RB-RB vs. II-II) as a between-subjects factor confirmed that participants in the RB-RB group reached criterion significantly faster than participants in the II-II group, *t*(13.48) = -2.40, *p* = .031.

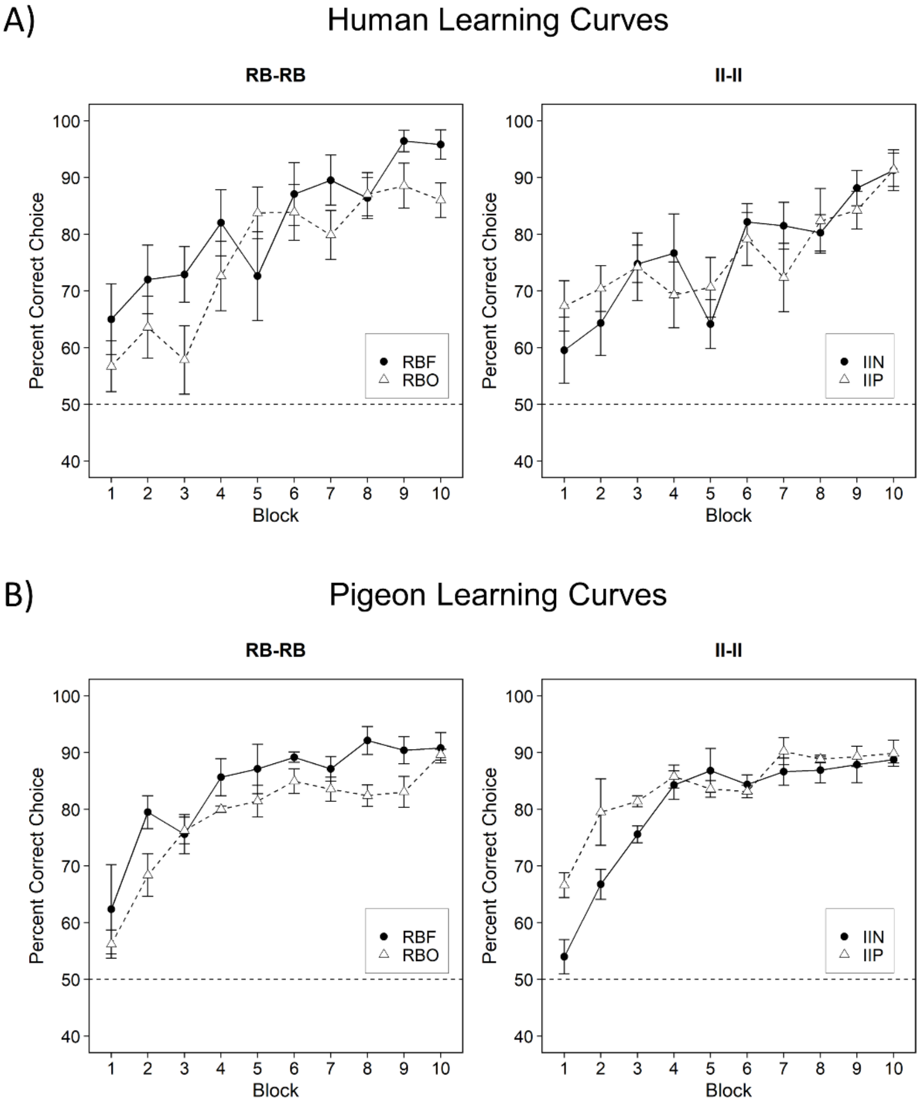

Figure 3. The top panels (A) depict learning curves for the humans in Experiment 1; the bottom panels (B) depict learning curves for the pigeons in Experiment 2A. In each case, individual participants' data were divided into 10 blocks (humans: mean block size = 6.60 trials, SD = 3.89 trials; pigeons: mean block size = 108.00 trials, SD = 74.81 trials). Error bars represent the standard error of the mean.

Decision Bound Modeling. Researchers investigating performance in RB and II paradigms typically implement manipulation checks to confirm that most RB and II learners arrived at the optimal strategies for each task condition (declarative rule use in the case of RB tasks; associative learning in the case of II tasks). The most commonly deployed manipulation check is decision bound modelling (DBM), a simplified version of General Recognition Theory (Ashby & Gott, 1988; Ashby & Perrin, 1988), which assumes that participants assign category membership in accordance with deterministic decision bounds placed in psychological stimulus space. In DBM, multiple models are fit to each participant's data to characterize their decision bound, and the best-fitting model is commonly assumed to reflect that participant's underlying strategy.

Typically, participants whose responses are best characterized by unidimensional decision bounds (i.e., decision bounds that lie perpendicular to the relevant dimension) are assumed to have relied on declarative rules, whereas participants whose responses are best characterized by diagonal decision bounds are assumed to have relied on associative mechanisms (e.g., Casale, Roeder, & Ashby, 2012; Maddox & Ashby, 2004; O'Donoghue, Broschard, Freeman, & Wasserman, 2022; Smith, Boomer, Zakrzewski, Roeder, Church, & Ashby, 2014; c.f., Edmunds, Milton, & Wills, 2018, and note that we revisit this assumption in Experiment 2).

In our DBM analysis, we initially[4] considered the four models that are most commonly used in the broader dual-systems literature: random guessing (RGM), which assumes that participants respond at random (allowing for possible response bias); two one-dimensional models, each of which assumes that participants used a single stimulus dimension to guide their responding (either spatial frequency in the case of the 1DF model or line orientation in the case of the 1DO model); and a linear two-dimensional model (2DL), which assumes that participants used both spatial frequency and line orientation. Assuming that most learners converge on the optimal strategies for their respective task conditions, most successful RB-RB learners' choice behavior should be best-described by a one-dimensional model (1DF in the RBF subtask; 1DO in the RBO subtask), whereas most

---

[4] One valid critique of much existing RB-II literature is that using just these four models may be insufficient to capture the full range of strategies that participants might actually deploy to solve RB and II tasks. Of particular note, Edmunds et al. (2018) demonstrated that participants who use conjunctive rules to reach criterion on II tasks are often misidentified as associative learners (because they are best fit by linear two-dimensional models) *even when* conjunctive rule-based models are also fit to their data. With this result in mind, and for parsimony with the existing RB-II literature, we focus on just four models here; but, in the Supplemental Materials, we detail an alternative DBM analysis that also includes a fifth, conjunctive rule-based model. Additionally, we revisit broader theoretical concerns surrounding DBM in the *General Discussion*.

successful II-II learners' choice behavior should be best-described by the 2DL model in both the IIN and IIP subtasks.

For each participant, and for each subtask on which that participant was trained, we fit the four models to the last 30 trials (i.e., to the 85% criterion window) from that participant's training phase. Through supplementary analyses, we also ruled out the possibility of strong idiosyncrasies within these 30-trial windows (which might, for example, have made some participants' II distributions more amenable to rule use than others'; see Supplementary Materials). We used the Bayesian Information Criterion (BIC; Neath & Cavanaugh, 2012) as our measure of model fit. The BIC penalizes more complex models (i.e., models with more free parameters) more heavily, and a lower BIC indicates a better fit to the data. All model fits were calculated using the 'grt' package for R (Matsuki, 2017).

The first subsection of Table 1 ("Humans (Experiment 1) – Individual Subtask Training") summarizes the aggregate results of our DBM analysis for the Individual Subtask Training phase (see Supplemental Materials for details of individual participants' model fits). Overall, and as expected, most participants in each subtask were best-described by the optimal model for that subtask. In the RB-RB group, 13 of 15 participants (86.67%) were best-described by the 1DF model in the RBF subtask, and 11 of 15 participants (73.33%) were best-described by the 1DO model in the RBO subtask. Fisher's Exact Tests revealed that neither distribution significantly diverged from the optimally expected distributions (wherein 100% of participants would converge on the optimal unidimensional model for each subtask), both *p*s > 0.996.

Table 1.

Subtask	*RGM*	*1DF*	*1DO*	*2DL*
Humans (Experiment 1) – Individual Subtask Training				
RBF	–	86.67% (13)	–	13.33% (2)
RBO	–	–	73.33% (11)	26.67% (4)
IIN	–	27.27% (3)	18.18% (2)	54.54% (6)
IIP	–	27.27% (3)	9.09% (1)	63.64% (7)
Humans (Experiment 1) – Task Switching				

RBF	–	86.67% (13)	–	13.33% (2)
RBO	6.67% (1)	–	60.00% (9)	33.33% (5)
IIN	–	27.27% (3)	18.18% (2)	54.54% (6)
IIP	–	36.36% (4)	18.18% (2)	45.45% (5)
Pigeons (Experiment 2A) – Individual Subtask Training				
RBF	–	75.00% (3)	–	25.00% (1)
RBO	–	–	75.00% (3)	25.00% (1)
IIN	–	–	–	100.00% (4)
IIP	–	–	–	100.00% (4)
Pigeons (Experiment 2A) – Task Switching				
RBF	–	75.00% (3)	–	25.00% (1)
RBO	–	–	75.00% (3)	25.00% (1)
IIN	–	–	–	100.00% (4)
IIP	–	–	–	100.00% (4)

Table 1. DBM model fits. For each of the humans in Experiment 1 (first four rows) and the pigeons in Experiment 2A (last four rows), the cell values indicate the percentage of participants in each task condition who were best-described by each model, with the absolute number of participants in parentheses.

The model fits for participants in the II-II group were slightly more variable (although the variability observed here is comparable to that observed in past research; e.g., Casale et al., 2012; Edmunds, Milton, & Wills, 2015; Edmunds et al., 2016). Nevertheless, and as expected, the majority of participants were best described by the 2DL model. Six of the 11 participants (54.54%) were best described by the 2DL model in the IIN subtask, and seven of the 11 participants (63.64%) were best described by the 2DL model in the IIP subtask. In the broader RB-II literature, these findings would commonly be taken to suggest that most participants converged on the optimal strategies for each task condition (c.f., Edmunds et al., 2016; Edmunds et al., 2018). Nevertheless, Fisher’s Exact Tests revealed that, in both the IIN and IIP subtasks, our data did diverge from the optimally expected distributions (wherein 100% of participants would be expected to converge on the 2DL model), both *ps* < .022.

Self-Report. As an additional indicator of which mechanisms participants might have used in each of the RB-RB and II-II conditions, we next explored our own participants’ descriptions of how they solved each subtask. Formal dual-systems models of categorization typically make no predictions surrounding self-report (e.g., Ashby et al., 1998), and implicit theorists often assume that self-reports cannot yield relevant insights into associative processes (e.g., Ashby & Rosedahl, 2017). Nevertheless, some researchers have asserted that participants' self-reports *could* plausibly differ between RB and II task conditions (e.g., Edmunds et al., 2016), while still others have reported that they *do* differ (e.g., Chandrasekaran, Koslov, & Maddox, 2014). With these opposing perspectives in mind, we were interested in exploring the concordance between self-reported strategy and other potential indicators of participants' learning mechanisms (e.g., DBM; see also Edmunds et al., 2016).

As described in the Method, participants were prompted to self-report which approaches they took to categorizing the stimuli after completing each of Phases 1 and 2 (Individual Subtask Training). All responses were independently coded by three raters (HF, RB, and EO) according to the following criteria, which were initially based on the guidelines developed by Edmunds et al. (2016) and then modified according to pilot data from a related project conducted in our laboratory. All raters were blind as to which task conditions and subtasks corresponded to which self-reports.

Participants were classified as *unidimensional rule users* if they reported using a single stimulus dimension (e.g., "the direction of the stripes")[5]. We additionally subdivided unidimensional rule users according to whether they relied on spatial frequency ("Freq") or line orientation ("Orient").

Participants were classified as *bidimensional rule users* if they reported using both stimulus dimensions in an explicitly rule-like manner. This distinction includes participants who developed a unidimensional rule with a small number of exceptions ("RuleEX"; e.g., "it depended on the angle of the stripes, but if bars were thick enough, then the correct response was always 'q'"), participants who described a clear strategy based on distance from the prototype of one category ("A/not A"; e.g., "the correct response was 'q' if the angle was close to 90° and if the lines were thin enough,

---

[5] The sample self-reports provided throughout this section are illustrative of our coding scheme and do not represent real data.

otherwise it was ‘p’”), and participants who developed any other verbalizable rule based on both stimulus dimensions (“Misc.”; e.g., “thick, vertical lines were in one category; thin, horizontal lines were in the other category”).

Participants were classified as *implicit learners* if they reported relying on intuition, muscle memory, family resemblance, and/or having a ‘feeling’ for each response (“Implicit”).

Participants were classified as *miscellaneous strategy users* if they reported any of the following: using both spatial frequency and line orientation in an unspecified manner (“Unclear”; e.g., “it depends on the width and the angle of the bars”, which could plausibly constitute either unspecified rule use or similarity-based, associative classification); guessing randomly (“Guessing”); explicitly memorizing the correct stimulus-response assignments (“Memory”)[6], or relying on aspects of the task that were irrelevant to category membership (“Irrel.”; e.g., “I think there was a set sequence of correct responses”).

Table 2 depicts the raters’ classifications of our participants’ self-reports. Inter-rater reliability was strong, Fleiss’ κ = .83. In cases of disagreement, we selected the strategy that was agreed upon by the majority of the raters (there were no instances in which all three raters selected different strategies).

Table 2.

	Unidimensional Rules		*Bidimensional Rules*			*Implicit*	*Miscellaneous Strategies*			
Subtask	*Freq*	*Orient*	*RuleEX*	*A/not A*	*Misc.*	*Implicit*	*Unclear*	*Guessing*	*Memory*	*Irrel*
RBF	93.33% (14)	–	–	–	–	–	6.67% (1)	–	–	–
RBO	–	100.00% (15)	–	–	–	–	–	–	–	–
IIN	36.36% (4)	9.09% (1)	9.09% (1)	–	–	–	45.45% (5)	–	–	–
IIP	36.36% (4)	–	–	–	18.18% (2)	–	45.45% (5)	–	–	–

---

[6] In our tasks, the stimuli were always randomly sampled; the likelihood of any individual stimulus appearing twice was effectively zero. However, in related pilot projects, a small subset of participants reported believing that the stimuli *were* repeated across trials and that the correct category-response assignments could be memorized.

Table 2. Self-reported strategies in Experiment 1. The cell values indicate the percentage of participants in each task condition who were classified as having reported each strategy, with absolute values in parentheses.

The vast majority of participants in the RB-RB group reported using the optimal strategies to solve each subtask: 93.33% of participants trained on the RBF subtask reported using a unidimensional rule based on spatial frequency, and 100.00% of participants trained on the RBO subtask reported using a unidimensional rule based on line orientation. Neither of these distributions significantly diverged from the optimally expected distribution (wherein 100% of participants would report using the optimal unidimensional rule), both Fisher's Exact *p*s = 1.00.

The participants in the II-II group displayed much more variability in their reports. In each of the IIN and IIP subtasks, a slim majority of participants (54.54%) reported using either a unidimensional or a bidimensional decision rule. In theory, a single unidimensional decision bound could not achieve the 85% criterion on either II subtask (Figure 1B; and see Supplemental Materials for evidence that there were no strong idiosyncrasies across individual participants' II distributions). However, multiple participants noted that, although they believed that the task was solvable by a unidimensional rule, they found it difficult to determine the exact category boundary. This uncertainty, in conjunction with our random stimulus sampling, may have allowed a small subset of unidimensional rule users to reach the 85% criterion.

The remainder of participants (45.45%) reported using both spatial frequency and line orientation but did not provide further information as to how those dimensions were used, meaning that we could not confidently classify them as either declarative rule users or associative learners.

Interestingly, *no* participants in the II-II group were classified as implicit learners according to our criteria; indeed, the observed distributions for each subtask significantly differed from the optimally expected distributions, wherein 100% of participants would report using Implicit strategies, both Fisher's Exact *p*s < .001. Of course, we acknowledge that self-reports may be inaccurate, and that asking participants to verbally describe their strategies might have biased them toward generating verbalizable decision rules (even if they did not actually use those rules to achieve task mastery). With this consideration in mind, we also acknowledge that at least some participants who were classified as miscellaneous bidimensional learners (“Unclear”; Table 2) may have deployed associative learning mechanisms. These participants did report *which* dimensions were relevant, but they may have been unable to describe *how* those dimensions were relevant, which would be consistent with the suggestion that associative learners might have difficulty accessing their memory representations (Edmunds et al., 2016; Smith et al., 2012). Nevertheless, even if we were to assume that *all* II-II group participants classified as having ‘Unclear’ self-reports were actually associative learners, then associative learners would still constitute less than half of our sample.

Task switching. Finally, we considered the data from Phase 3 (Task Switching). Importantly, participants’ strategies could conceivably have changed between Phases 1 and 2 (Individual Subtask Training) and Phase 3 (Task Switching). For example, the enhanced cognitive load imposed in Phase 3 might have encouraged participants to adopt simpler strategies than they had previously used in the initial training phases (resulting in higher proportions of random guessing in RB tasks, and/or higher proportions of unidimensional rule use in II tasks).

To address this possibility, we used DBM to describe each participant’s best-fitting decision bounds at the end of Phase 3 (Task Switching). We separately considered the final 100 trials from each of the two subtasks on which each individual participant had been trained (RBF and RBO in the RB-RB condition; IIN and IIP in the II-II condition). As evidenced by Table 1, participants’ best fitting decision bounds remained broadly consistent between Phases 1 and 2 (Individual Subtask Training) and Phase 3 (Task Switching). Indeed, the group-level distributions for each of the RBF and IIN subtasks were identical across phases, while the group-level distributions for each of the RBO and IIP subtasks displayed only slight differences (these differences were nonsignificant in each case, both

Fisher's Exact *p*s > .700). Likewise, in our post-experiment questionnaire (see Method), no participant reported changes in strategy during the task switching phase.

Having determined that participants' strategies did not meaningfully differ as a consequence of task switching, we next considered their task switching performance. Given the dissociations that have previously been reported between declarative and associative task switching paradigms (e.g., Forrest et al., 2012; Meier et al., 2013; Li et al., 2019a), we expected to observe large switch costs and small congruency effects in RB-RB group, whereas we expected to observe small switch costs and large congruency effects in the II-II group.

Overall, performance in the task switching phase was robust (Figure 4A). Participants in the RB-RB group achieved an average accuracy of 84.74% (SD = 7.03%), whereas participants in the II-II group achieved an average accuracy of 70.21% (SD = 15.26%). To assess participants' task switching performance, we classified each trial as either a *switch trial* (meaning that trial *N* involved a different subtask than trial *N-1*) or a *stay trial* (meaning that trial *N* shared the same subtask as trial *N-1*). Additionally, we classified each stimulus as either *congruent* (meaning that it shared the same correct response under both trained subtasks) or *incongruent* (meaning that the correct response differed depending on the subtask).

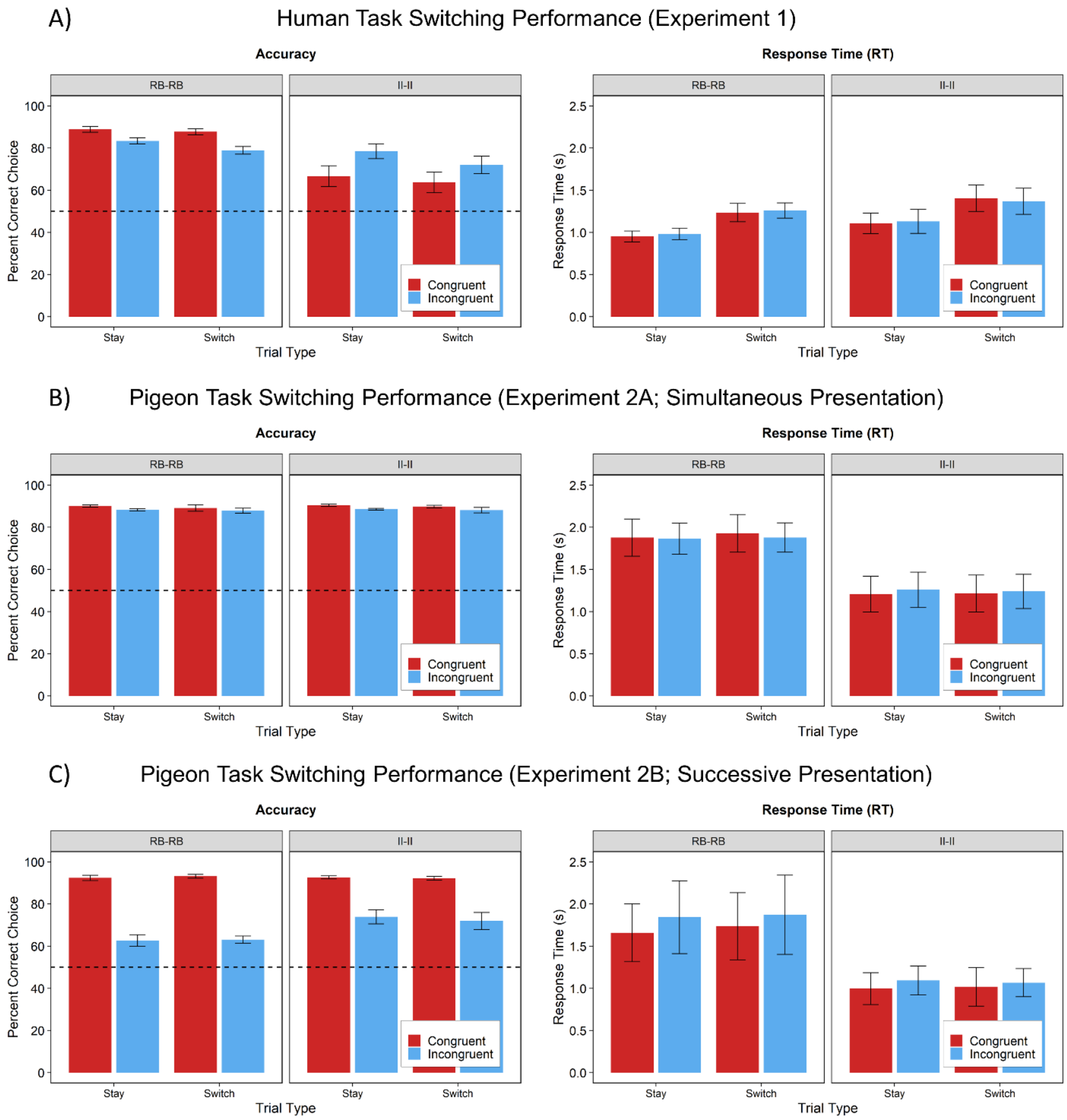

Figure 4. Task switching performance. The top panels (A) depict humans' performance in Experiment 1; the middle panels (B) depict pigeons' performance in Experiment 2A; the bottom panels (C) depict pigeons' performance in Experiment 2B. In each case, the lefthand panels depict participants' choice accuracies, and the righthand panels depict participants' response times. Error bars reflect the standard error of the mean, and the dashed lines in the lefthand panels represent chance-level performance.

Next, we asked whether participants' choice accuracies varied as a function of trial type and/or congruency. We submitted participants' choice accuracies to a logistic regression with task condition (RB-RB vs. II-II; effects-coded), trial type (stay vs. switch; dummy-coded with stay trials as the reference), and response congruency (congruent vs. incongruent; dummy-coded with congruent trials as the reference) as factors.[7]

The main effect of task condition was significant, *B* = -0.69, *SE* = 0.04, *Z* = -16.93, *p* < .001; however, it was qualified by a significant interaction between task condition and congruency, *B* = 0.53, *SE* = 0.05, *Z* = 9.24, *p* < .001. To probe this interaction, we conducted two follow-up logistic regressions that separately examined the effect of congruency for each of the RB-RB and II-II task conditions. The effect of congruency was significant in each case; however, whereas participants in the RB-RB group were *less* accurate on incongruent trials (mean = 81.17%, SD = 6.74%) than on congruent trials (mean = 88.31%, SD = 5.36%), *B* = -0.53, *SE* = 0.07, , *Z* = -7.78, *p* < .001, participants in the II-II group were *more* accurate on incongruent trials (mean = 76.11%, SD = 12.22%) than on congruent trials (mean = 65.97%, SD = 14.98%), *B* = 0.52, *SE* = 0.06, *Z* = 8.46, *p* < .001.

The fact that participants in the II-II group performed better on incongruent trials is surprising; however, this pattern may reflect the way in the stimulus distributions were rotated to generate each subtask. In the II-II group, the incongruent regions of the stimulus space encompassed the extreme ends of the frequency dimension, whereas in the RB-RB group, the extreme ends of the frequency distribution were 50% congruent and 50% incongruent (see Supplemental Materials). Past research in our laboratory has found that spatial frequency may be more salient than line orientation both for people and pigeons (e.g., O’Donoghue et al., 2020; O’Donoghue et al., 2022; see also Herbranson, Karas, & Hardin, 2017); so, among participants in the II-II group, the benefits of extreme

---

[7] For all regression models reported in these experiments, we began by calculating the intraclass correlation coefficient (ICC) for an intercept-only model to determine whether mixed-effects modelling was justified. We used the guideline that an ICC < .10 indicates no advantage of mixed-effects modeling over standard regression, whereas an ICC >= .10 justifies mixed-effects modeling (García-Patos & Olmos, 2020). Ultimately, *none* of our accuracy analyses incorporated mixed-effects modeling, whereas *all* of our reaction time analyses incorporated random intercepts at the subject level. Although not reported here, we also confirmed that none of our accuracy analyses meaningfully differed when random intercepts were included. In our mixed-effects models, all *t*-tests were calculated using the lmerTest package for R Studio (Kuznetsova, Brockhoff, & Christensen, 2017), which uses Satterthwaite's method to approximate degrees of freedom.

spatial frequency values may have outweighed any potential costs associated with incongruency. We are thus reluctant to draw strong conclusions based on our obtained congruency effects.

We also observed a significant interaction between trial type and congruency, *B* = -0.19, *SE* = 0.10, *Z* = -1.98*, p* = .048. To probe this interaction, we separately examined the effect of trial type at each level of congruency. For congruent stimuli, the effect of trial type was nonsignificant, *B* = -.09, *SE* = 0.07, *Z* = -1.43, *p* = .153; but, for incongruent stimuli, accuracy was significantly lower on switch trials (mean = 75.99%, SD = 10.69%) than on stay trials (mean = 81.33%, SD = 8.99%), *B* = -0.31, *SE =* 0.06, *Z* = -4.75, *p* < .001. Similar interactions have previously been reported in the task switching literature [e.g., Forrest et al., 2012; Gopher, Armony, & Greenshpan, 2000 (Experiment 1); Hughes, Linck, Bowles, Koeth, & Bunting, 2014], and they may be driven by the fact that, when responding to congruent stimuli, participants need not attend to the task cue.

No other main effects or interactions were significant, all *p*s > .092. The absence of a main effect of trial type (and of higher-order interactions between task condition and trial type) was surprising given that accuracy-based switch costs have been observed in past literature (e.g., Meier et al., 2013; Monsell, 2003). However, speed-accuracy trade-offs can also preclude differences in accuracy while still yielding strong switch costs in response times (Samavatyan & Leth-Steensen, 2009; see also Li et al., 2019b). So, we next considered participants’ response times. We logged all response times for analysis, and excluded all incorrect trials as well as all response times that fell more than three standard deviations beyond each individual participant’s mean score. We then conducted a linear mixed-effects model analysis (see footnote 4) with random intercepts at the subject-level, and with task condition (RB-RB vs II-II), congruency (congruent vs. incongruent), and trial type (stay vs. switch) as factors, all coded as previously described.

As above, we observed a significant interaction between task condition and congruency, *b* = -0.01, *SE* < 0.01, *t*(9678) = -2.40, p = .016. Follow-up probes indicated that participants in the RB-RB group were slower to respond to incongruent stimuli (mean = 1120.84 ms, SD = 334.20 ms) than to congruent stimuli (mean = 1092.87 ms, SD = 370.81 ms), *b* = 0.01, *SE* < .01, *t*(6011) = 3.85, p < .001, whereas participants in the II-II group displayed no differences in response time as a function of congruency, *b* < .01, *SE* < .01, *t*(3670) = 0.40, *p* = .687.

Critically, we also observed a main effect of switch type, *b* = 0.09, *SE* < 0.01, *t*(9676) = 17.07, *p* < .001, indicating that participants were generally slower to respond on switch trials (mean = 1305.26 ms, SD = 446.15 ms) than on stay trials (mean = 1035.07 ms, SD = 349.23 ms). No other main effects or interactions were significant, all *p*s > .141.

Of particular note, the fact that the two-way interaction between task condition and trial type was not significant indicates that, contrary to our predictions, switch cost magnitudes did not meaningfully differ between the RB-RB and II-II task conditions. The fact that both the RB-RB group *and* the II-II group displayed comparable switch costs suggests two interpretive possibilities. First, participants may have deployed declarative learning mechanisms regardless of their assigned task condition; that is, in our paradigm, the majority of successful II learners may have relied on declarative decision rules rather than on associative learning mechanisms (see also Edmunds et al., 2016). This conclusion would be consistent with participants' self-reported strategies, but broadly inconsistent with the results of DBM, and we revisit it (alongside the broader debate surrounding whether RB and II tasks reliably encourage different learning mechanisms) in the *General Discussion.*

Alternatively, it is possible that switch costs might not reliably dissociate declarative rule users from associative learners in the context of tasks that preclude exemplar memorization. To further explore this possibility, we turned to a new participant sample – and to a new species – in Experiment 2.

Experiment 2A

In Experiment 2A, we asked whether pigeons – which appear to rely solely on associative learning mechanisms to solve RB and II tasks (O'Donoghue et al., 2020; Qadri et al., 2019; Smith et al., 2012), as well as various other paradigms that encourage rule-use among human participants (Lea, Wills, Leaver, Ryan, Bryant, & Miller, 2009; Maes et al., 2015; Navarro, Jani, & Wasserman, 2019) – would display switch costs when trained to solve RB and II tasks in the same manner as the human participants in Experiment 1. Based on past comparative research (Castro & Wasserman, 2016; Meier et al., 2013; Meier et al., 2016; O'Donoghue & Wasserman, 2021), we expected that pigeons would not display significant switch costs in either task condition (RB-RB or II-II), consistent with the possibility that associative learning mechanisms can preclude switch costs.

Method

Subjects and Apparatus. Eight pigeons took part in Experiment 2A. Our sample size was determined by availability; we selected only the pigeons in our laboratory that had no prior experience with the present tasks or stimuli. The pigeons were individually housed and maintained at 85% of their free-feeding weights. All experimental procedures were approved by the Institutional Animal Care and Use Committee at The University of Iowa.

The pigeons were trained in 36 x 36 x 41 cm operant conditioning chambers with white noise played during all sessions. The stimuli were presented on 800 x 600 px LCD monitors equipped with AccuTouch® touchscreens (Elo TouchSystems, Fremont, CA). The screen area visible to each pigeon measured 28.5 x 17 cm. Food reinforcement (1 to 3 45-mg pigeon pellets) was delivered on each trial via a rotary dispenser connected to a food tray mounted on the wall opposite the touchscreen.

Stimuli and Apparatus. We programmed Experiment 2A and generated the stimuli using Matlab and its Psychtoolbox-3 extensions (Kleiner, Brainard, & Pelli, 2007), using the same stimulus sampling procedures as in Experiment 1.

Phases 1 and 2: Individual Subtask Training. As were the humans in Experiment 1, the pigeons were trained to criterion on each individual subtask prior to entering the critical task switching phase. Four pigeons were randomly assigned to each of the RB-RB and II-II conditions. The order in which the subtasks were trained was counterbalanced across the pigeons in each condition.

Each trial began with the presentation of a 4.50 x 4.50 cm star-shaped start stimulus in the center of the screen (Figure 2B). After the pigeon pecked the start stimulus once, the start stimulus disappeared and was replaced by a 7.00 x 7.00 cm circular task cue. As in Experiment 1, this task cue could be either blue (RGB value: [0, 0, 255]) or red (RGB value: [255, 0, 0]). Cue-subtask relationships were counterbalanced across pigeons. The pigeons had to peck the task cue once, at which point the 4.50 x 4.50 cm trial stimulus was overlaid on the cue. The pigeons then had to complete a variable response requirement to the trial stimulus, which began with one peck and was adjusted upwards with performance to a maximum of 10 pecks (such upward adjustment is common in the comparative literature because larger response requirements promote attention to onscreen stimuli, while the gradual increase helps ensure that the pigeons do not suddenly stop responding altogether; e.g., Kelleher, Riddle, & Cook, 1962). Once the response requirement was completed, the trial exemplar disappeared and was immediately replaced with the two choice keys, which were replicas of the trial exemplar that appeared to the left and right sides of the screen, 4.00 cm from the nearest edge. Each choice key was overlaid on a replica of the 7.00 x 7.00 cm circular task cue. Choice key-category relationships (e.g., [left key-Category A]; [right key-Category B]) were counterbalanced across pigeons, and a single peck to either choice key was recorded as a response. Correct responses were followed by food reinforcement, whereas incorrect responses were followed by a variable 6-10 s delay, and then by a correction trial. Correction trials continued until the pigeons made the correct response. The next trial then began after a variable 6-10 s intertrial interval (ITI).

Phase 3: Blocked Subtask Training. Because the pigeons were trained over a much longer timeframe than the humans in Experiment 1, we worried that they might have forgotten some aspects of the first subtask (Phase 1) after reaching criterion on the second subtask (Phase 2). So, before advancing the pigeons to the critical task switching phase, we next shifted them onto a blocked training phase wherein both subtasks were presented in a blocked sequence within each individual session. Each subtask continued for 60 trials, with each full session comprising 120 trials. The procedure was otherwise identical to that described above.

We had originally planned to train the pigeons to an 85% criterion in the blocked subtask training phase; however, this criterion proved slightly too challenging for most pigeons to achieve (note that this phase of training functionally comprised a partial midsession reversal – a procedure wherein 50% of the prevailing reward contingencies are reversed midway through an experimental session – which is very difficult for pigeons to master; Zentall, 2020). We instead trained each pigeon for a fixed number of sessions (30 sessions) before advancing that pigeon to the intermixed task switching phase. Because the blocked subtask training phase was not of theoretical interest, we did not further analyze the data from it.

Phase 4: Task switching. Finally, each pigeon progressed to the task switching phase. As for the humans in Experiment 1, the subtasks were now intermixed, with the trial order partially randomized such that a subtask switch occurred on one third of all trials. The trial sequence was otherwise identical to that described above. We maintained each pigeon on the task switching phase for 24 sessions, with each session comprising 120 trials.

Results and Discussion

Learning Speeds. Figure 3B depicts our pigeons' learning curves throughout Phases 1 and 2 (Individual Subtask Training). The pigeons reached the 85% learning criterion rapidly in each of the RBF (mean = 5.00 sessions, min = 4.00 sessions, max = 7.00 sessions), IIN (mean = 6.25 sessions, min = 3.00 sessions, max = 10.00 sessions), and IIP subtasks (mean = 6.25 sessions, min = 5.00 sessions, max = 9.00 sessions). Reaching criterion in the RBO subtask proved to be more difficult (mean = 18.50 sessions, min = 11.00 sessions, max = 24.00 sessions). This finding is consistent with prior research in our laboratory, as well as with the broader literature suggesting that, for pigeons, spatial frequency is more salient than line orientation (Herbranson et al., 2017; O'Donoghue et al., 2020).

After collapsing across the subtasks in each task condition (RBF and RBO; IIN and IIP), an independent-samples *t*-test revealed that, on average, the pigeons in the RB-RB group took significantly *longer* to reach criterion than the pigeons in the II-II group, $t(5.66) = 2.99$, $p = .026$. Given that pigeons do not typically display overall learning speed differences between RB and II paradigms (O'Donoghue et al., 2020; Smith et al., 2012), this finding was unexpected. Nevertheless, the difference was driven by the amount of time required to reach criterion in the RBO subtask and, critically, the direction of the effect is inconsistent with the possibility that the pigeons may have deployed declarative decision rules (in which case the pigeons in the RB-RB group should have reached criterion more quickly than the pigeons in the II-II group).

Decision Bound Modeling. As in Experiment 1, we next used decision bound modeling (DBM) to characterize the pigeons' decision bounds in Phases 1 and 2 (Individual Subtask Training). We separately fit the same four models described in Experiment 1 (RGM, 1DF, 1DO, and 2DL) to the final 240 trials (i.e., to the 85% criterion window) from each pigeon, and from each subtask on which that pigeon was trained.[8]

As detailed in Table 1, most pigeons were best described by the optimal decision bound for each task condition. In the RB-RB group, three of the four pigeons (75%) were best-described by the 1DF model in the RBF subtask; likewise, three of the four pigeons (75%) were best-described by the 1DO model in the RBO subtask. In the II-II group, all four pigeons (100%) were best-described by the 2DL model in each of the IIN and IIP subtasks. Across all four subtasks, our observed distributions did not differ significantly from the optimally expected distributions (wherein 100% of pigeons would converge on the optimal model), all Fisher's Exact *p*s = 1.00.

---

[8] Note that we modelled 240 trials from each pigeon in Experiment 2A, but just 30 trials from each human in Experiment 1. We selected different window sizes because we wanted to assess both species' performance at a common accuracy criterion (85%); however, we acknowledge that including different numbers of trials might impact model fits. For transparency, we also ran a separate DBM analysis using just the last 30 trials from each individual pigeon, and from each subtask on which that pigeon was trained. The results do not differ from those reported here except that, with a 30 trial window, 100% of the pigeons trained on the RBO subtask were best described by the 1DO model.

If we interpreted these data as humans' data are typically interpreted (Experiment 1; see also Casale et al., 2012; O'Donoghue et al., 2022; Smith et al., 2014), then we might be tempted to conclude that most pigeons in the RB-RB group deployed declarative decision rules. This claim would surely be controversial. The between-group differences in learning speeds that we observed in Experiment 2 were inconsistent with this possibility, and to our knowledge, pigeons have thus far shown no compelling evidence for declarative rule use in *any* categorization paradigm that promotes rule use among humans (Lea et al., 2009; Maes et al., 2015; Navarro et al., 2019; O'Donoghue et al., 2020; Smith et al., 2012).

However – and critically – DBM is not a process model (Ashby & Valentin, 2018). Any individual decision bound is compatible with multiple underlying processes, meaning that a participant's best-fitting decision bound may not always be useful for inferring which mechanisms that participant actually deployed. Here, the RB-RB group's model fits could readily be accommodated by the fact that unidimensional sorting need not implicate declarative rule use (Wills et al., 2009; see also Smith et al., 2011). We agree with this interpretation, but also stress that the same caveats and considerations must apply when interpreting humans' model fits (Edmunds et al., 2016; Edmunds et al., 2018). We revisit this issue in the General Discussion.

Task switching. Finally, we considered the pigeons' task switching performance (Figure 4B). As in Experiment 1, we first sought to confirm that the pigeons' response strategies (as indexed by DBM) did not meaningfully differ between Phases 1 and 2 (Individual Subtask Training) and Phase 4 (Task Switching). To do so, we fit DBM separately to the final 100 trials from each individual subtask that each pigeon was exposed to in the task switching phase (RBF and RBO in the RB-RB group; IIN and IIP in the II-II group). The group-level distributions of best-fitting models were identical across phases for all four subtasks (Table 1), indicating that the introduction of the task switching requirement did not meaningfully influence pigeons' decision bounds.

As had the humans in Experiment 1, the pigeons achieved noteworthy levels of accuracy in the task switching phase (RB-RB group: mean = 88.83%, SD = 2.08%; II-II group: mean = 89.94%, SD = 1.72%). We assessed the pigeons’ task switching performance using the same approach and the same statistical models described in Experiment 1 [briefly: we considered trial type (stay vs. switch), congruency (congruent vs. incongruent), and task condition (RB-RB vs II-II) as predictors of both choice accuracies and response times]. In our analysis of choice accuracy, we observed a significant main effect of congruency, *B* = -0.19, *SE* = 0.05, *Z* = -3.62, *p* < .001, indicating that the pigeons were generally less accurate on incongruent trials (mean = 88.21%, SD = 1.73%) than on congruent trials (mean = 89.86%, SD = 1.71%). No other main effects or interactions were significant, all *p*s > 0.214.

In our analysis of response times, we likewise obtained a significant main effect of congruency, *b* = 0.01, *SE* < 0.01, *t*(20010) = 2.34, *p* = .019; however, this effect was qualified by a significant interaction between task condition and congruency, *b* = 0.01, *SE* < 0.01, *t*(20010) = 2.65, *p* = .008. No other main effects or interactions were significant, all *p*s > .050. Follow-up probes indicated that, for the pigeons in the RB-RB group, response times did not reliably differ as a function of congruency, *b* = -.01, *SE* < .01, *t*(9982) = -1.77, *p* = .077. Conversely, the pigeons in the II-II group were significantly slower to respond on incongruent trials (mean = 1249.04 ms, SD = 380.69 ms) than on congruent trials (mean = 1210.55 ms, SD = 400.21 ms), *b* = .02, *SE* < .01, *t*(10030) = 3.87, *p* < .001.

Importantly, the pigeons did not display any evidence of switch costs in either their accuracies or their response times. This finding is consistent with past research involving task switching among pigeons (Castro & Wasserman, 2016; Meier et al., 2013; Meier et al., 2016; O’Donoghue & Wasserman, 2021) and with our original assumption that ‘true’ associative learners may not display strong switch costs in RB and II paradigms.

Nevertheless, recent research on pigeons’ task switching performance prompted us to consider an alternative possibility: namely, that the lack of switch costs displayed by our pigeons may *not* result from their deployment of associative learning mechanisms, but instead from the manner in which they process the events of each trial.

Meier and colleagues (Meier et al., 2013; Meier et al., 2016) have previously argued that, when task cues and trial exemplars are simultaneously visible (as they were in Experiments 1 and 2A; Figure 2), humans may generally process those trial events in a hierarchical manner, whereas pigeons may generally process those trial events in a configural manner. Indeed, we recently found support for configural processing of task cues and discriminative stimuli among pigeons trained on a simultaneous conditional discrimination (O’Donoghue, Castro, & Wasserman, 2022).

Critically, whereas hierarchical processing (among humans) may produce switch costs, configural processing (among pigeons) may preclude switch costs. As argued by Meier et al. (2013), configural processing may lead pigeons to perceive the same trial exemplar in a very different manner depending on which task cue accompanies it – meaning that, from a pigeon’s perspective, switch trials are functionally no different from stay trials. With this possibility in mind, in Experiment 2B, we asked whether pigeons would display switch costs under conditions that were quite *unlikely* to facilitate configural processing. If they did not, then we could be more confident that the lack of switch costs observed in Experiment 2A reflects the participation of associative learning mechanisms, rather than interspecies differences in stimulus processing.

Experiment 2B

Past research supports the conclusion that, whereas simultaneous stimulus presentation may be more likely to encourage configural processing, successive stimulus presentation may be more likely to encourage hierarchical stimulus processing (e.g., Holland, 1985). So, in Experiment 2B, we asked whether pigeons might produce switch costs when the task cue and the trial exemplar were *not* simultaneously available.

Method

Subjects, Apparatus, and Stimuli. We used the same subjects (*N* = 8), apparatus, and stimulus sampling procedures as in Experiment 2A. The pigeons progressed to Experiment 2B immediately after completing Phase 4 (task switching) in Experiment 2A.

Phase 1: 50% Successive, 50% Simultaneous. We worried that, if the pigeons had processed the task cues and trial exemplars configurally in Experiment 2A, then abruptly shifting them to entirely successive cue-exemplar presentation might disrupt their performance. Thus, we began with a preliminary phase of training wherein 50% of the trials involved simultaneous cue-exemplar presentation (as in Experiment 2A; Figure 2B), while the other 50% of trials involved successive cue-exemplar presentation (Figure 2C). On successive trials, the task cue disappeared as soon as the pigeon completed its observing response requirement; then, the trial exemplar appeared after a brief (10 ms) delay. The choice buttons were likewise presented without the task cues that had previously accompanied them in Experiment 2A (Figure 2).

The proportions of simultaneous and successive trials were evenly divided across the two subtasks on which each pigeon had been trained. As in the task switching phase of Experiment 2A, the subtasks themselves were pseudo-randomly intermixed such that a subtask-switch occurred on one third of all trials.

Each pigeon completed 30 sessions in this first phase of training, with each session comprising 120 trials (as in the blocked phase of Experiment 2A, we initially planned to train the pigeons to an 85% criterion; however, this criterion proved slightly too demanding for most pigeons to achieve). Because we were not theoretically interested in this phase of training, we did not further analyze the data from it.

Phase 2: 100% Successive. Finally, we shifted each pigeon to a second phase wherein *all* trials involved successive cue-exemplar presentation (Figure 2C). The trial sequence was otherwise identical to that described above. Each pigeon remained on this final phase for 24 sessions, with each session comprising 120 trials.

Results and Discussion

Task switching performance. Unsurprisingly, the successive cue-exemplar availability arranged here proved more challenging than the simultaneous cue-exemplar availability that had previously been arranged in Experiment 2A. Nevertheless, the pigeons achieved impressive levels of accuracy in the second, fully successive phase of training (Figure 4C; RB-RB group: mean = 77.82%, SD = 15.78%; II-II group: mean = 82.69%, SD = 11.19%).

We assessed task switching performance using the same general approach as in Experiment 2A. However, for all models, we conducted omnibus tests that combined the data from Experiments 2A and 2B; accordingly, each model also contained experimental phase (2A vs. 2B) as an additional predictor [alongside trial type (stay vs. switch), congruency (congruent vs. incongruent), and task condition (RB-RB vs. II-II)]. This omnibus approach revealed a main effect of experimental phase, confirming that accuracy was significantly lower in Experiment 2B relative to Experiment 2A, *B* = 0.29, SE = 0.06, Z = 4.96, p < .001.

In our omnibus analysis of choice accuracy, we additionally observed a significant main effect of congruency, *B* = 1.35, *SE* = 0.12, *Z* = 11.56, *p* < .001, which was qualified by significant two-way interactions between task condition and congruency, *B* = -0.25, *SE* = 0.12, *Z* = -2.14, *p* = 0.32, and between task condition and experimental phase, *B* = -1.54, *SE* = 0.07, *Z* = -21.18, *p* < .001, as well as by a significant three-way interaction between task condition, congruency, and experimental phase, *B* = 0.25, *SE* = 0.07, *Z* = 3.38, *p* < .001. No other main effects or interactions were significant, all *p*s > .196.

Follow-up probes to investigate the three-way interaction indicated that, in Experiment 2B, the pigeons displayed a two-way interaction between task and congruency that they had not previously displayed in Experiment 2A (Experiment 2A: *B* = 0.02, *SE* = 0.09, *Z* = 0.25, *p* = .803; Experiment 2B: *B* = -0.50, *SE* = 0.08, *Z* = -6.14, *p* < .001). The nature of this two-way interaction was such that, regardless of task condition, the pigeons were less accurate on incongruent trials than on congruent trials; however, this effect was stronger in the RB-RB task condition, *B* = -2.01, *SE* = 0.06, *Z* = -34.88, *p* < .001 (incongruent trials: mean = 62.87%, SD = 4.28%; congruent trials: mean = 92.78%, SD = 2.05%), than in the II-II task condition, *B* = -1.51, *SE* = 0.06, *Z* = -25.75, *p* < .001 (incongruent trials: mean = 72.93%, SD = 6.93%; congruent trials: mean = 92.44%, SD = 1.46%). As for the humans in Experiment 1, we hypothesize that the salience of extreme spatial frequency values (which entirely corresponded to incongruent stimuli for the pigeons in the II-II group, but not for the pigeons in the RB-RB group; see Supplemental Materials) may have provided a buffer against the effects of incongruency.

The two-way interaction between congruency and experimental phase also indicated that the congruency effects observed in Experiment 2B were significantly stronger than the congruency effects observed in Experiment 2A (Figure 4B; Figure 4C; Experiment 2A: *B* = -0.18, *SE* = 0.04, *Z* = -4.16, *p* < .001; Experiment 2B: *B* = -1.77, *SE* = 0.04, *Z* = -43.18, *p* < .001). This disparity likely reflects the fact that, under the successive contingencies arranged here, the pigeons needed to maintain the task cue in working memory in order to determine the correct response on incongruent trials. Conversely, on congruent trials, stimuli required the same response regardless of the task cue – so, the pigeons could respond correctly even if they had forgotten the task cue.

In our omnibus analysis of response times, we observed significant main effects of task condition, *b* =-0.10, *SE* = 0.04, *t*(38066) = -2.44, *p* = .049, and phase, *b* = -0.06, *SE* < .01 , *t*(38066) = -15.78, *p* < .001. These effects were qualified by significant two-way interactions between task condition and congruency, *b* = 0.03, *SE* < .01, *t*(38066) = 3.40, *p* < .001, and between experimental phase and congruency, *b* = 0.02, *SE* < .01, *t*(38066) = 2.91, *p* = .004, as well as by a significant three-way interaction between task condition, experimental phase, and congruency, *b* = -0.02, *SE* < .01, *t*(38066) = -3.37, *p* < .001. No other main effects or interactions were significant, all *p*s > .116.

Follow-up probes to investigate this three-way interaction indicated that, contrary to Experiment 2A – wherein the pigeons had previously displayed a two-way interaction between task condition and congruency, $b = 0.01$, $SE < .01$, $t(20023) = 3.99$, $p < .001$ – no such interaction was observed in Experiment 2B, $p = .092$. Instead, the pigeons in Experiment 2B displayed a simple main effect of congruency, $b = 0.02$, $SE < .01$, $t(18043) = 7.56$, $p < .001$, indicating that the pigeons were generally slower to respond on incongruent trials (mean = 1469.05 ms, SD = 731.38 ms) than on congruent trials (mean = 1352.39 ms, SD = 647.03 ms).

As in Experiment 2A, the pigeons displayed no evidence of switch costs despite the successive availability of the task cue and the trial exemplar, a manipulation that should encourage hierarchical stimulus processing (Holland, 1985). This finding indicates that the lack of switch costs observed in Experiments 2A and 2B likely reflects the contributions of pigeons' associative learning mechanisms, rather than interspecies differences in stimulus processing.

General Discussion

To date, most research involving task switching has investigated the contributions of declarative mechanisms, which reliably yield strong switch costs relative to weak congruency effects among human learners (Kiesel et al., 2007; Monsell, 2003). Conversely, among both human (e.g., Forrest et al., 2012; Li et al., 2019; Meier et al., 2013) and nonhuman learners (e.g., Castro & Wasserman, 2016; Meier et al., 2013; Meier et al., 2016; O'Donoghue & Wasserman, 2021), associative mechanisms may yield weak (or no) switch costs relative to strong congruency effects.

Thus far, investigations of humans' associative task switching performance have deployed small stimulus sets that facilitate (and at times explicitly encourage) exemplar memorization. Whether such memorization is truly associative remains a matter of debate (Edmunds et al., 2016; Bower & Winzenz, 1970; Forrest et al., 2012; Smith et al., 2012). In the present study, we asked whether participants would still display disparate task switching response profiles when trained to switch either between rule-based (RB) subtasks (which encourage declarative rule use) or between information-integration (II) subtasks (which encourage associative learning, but *discourage* exemplar memorization because they employ trial-unique stimuli).

In Experiment 1, we found that human participants displayed strong switch costs regardless of task condition. Conversely, in Experiment 2, we found that pigeons displayed no switch costs in either task condition. These conflicting results raise two possibilities: (1) that our human participants may have employed declarative mechanisms to solve both RB and II tasks, and (2) that at least among human learners, declarative and associative mechanisms may not reliably yield dissociable task switching performance. We discuss each of these considerations below.

Did Human II Learners use Declarative Rules?

Many researchers have argued that, whereas responding in RB tasks is predominately governed by declarative rule use, responding in II tasks is predominately governed by associative learning (see Ashby & Valentin, 2017, for a review). However, our human participants displayed task switching response profiles characteristic of declarative rule use in *both* RB *and* II paradigms, raising the possibility that, in Experiment 1, most successful II learners could have actually been rule-users (see also Edmunds et al., 2016; Edmunds et al., 2018; Nosofsky, Stanton, & Zaki, 2005; Wills et al., 2019; Zaki & Kleinschmidt, 2014).

The suggestion that several of our II-II group participants might have used declarative rules is consistent with their task switching response profiles and self-reported strategies, but inconsistent with the results of our DBM analyses, wherein most II-II group participants converged on the optimal two-dimensional model. Past research has likewise reported that self-report and DBM can yield opposing conclusions (e.g., Edmunds et al., 2016).

On the one hand, dual-systems perspectives often assume that self-reports cannot yield valid insights into implicit learning mechanisms (e.g., Ashby & Rosedahl, 2017). Conversely, some researchers have asserted that associative learners either conceivably might (e.g., Edmunds et al., 2016) or actually do produce self-reports reflecting their reliance on implicit processes; for example, Chandrasekaran et al. (2014) reported that many successful II learners reported responding according to “gut feeling”. Here, we make no strong claims concerning the validity of self-report; but, because self-reports can accurately predict performance in other category learning tasks (e.g., probabilistic categorization: Lagnado, Newell, Kahan, & Shanks, 2006), and because they have been used in efforts to capture various facets of implicit learning (e.g., Edmunds et al., 2016; Weinberger & Green, 2022), we do believe that they warrant further study in the context of RB and II learning.

In contrast to self-report, DBM provides an objective index of participants' actual categorization behavior. However, under at least some real-world conditions, DBM may not reliably allow for mechanistic inferences. A relevant consideration is raised by our data from Experiment 2A. In our DBM analysis, we found that most pigeons' data were best fit by the optimal models in each of the RB and II task conditions (unidimensional models in the case of RB tasks; bidimensional models in the case of II tasks). Based on this result, one *could* argue that – as did the humans in Experiment 1 – the pigeons in Experiment 2A used declarative rules to solve RB tasks. Given that pigeons have thus far shown no compelling evidence of rule use in *any* categorization paradigm known to promote rule use among humans (including RB paradigms, Smith et al., 2012; O'Donoghue et al., 2020; among others, Lea et al., 2009; Maes et al., 2015; Navarro et al., 2019), this interpretation is clearly implausible.

More reasonably, one could also conclude that the responses of pigeons in the RB-RB group were primarily controlled by the relevant stimulus dimension in each subtask (as indicated by their best-fitting decision bounds), but that such unidimensional control need not implicate declarative rule use (see Wills et al., 2019). We agree with this interpretation (see also Smith et al., 2011). However, we extend it further to stress that the same considerations must apply when interpreting humans' decision bounds. In particular, and although diagonal decision bounds among human II-learners do suggest attention to both dimensions, they need not implicate associative mechanisms (Edmunds et al., 2016; Edmunds et al., 2018; see also Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015). This consideration is particularly relevant because, of the four models that we considered (which we selected because they are the four models most commonly used in the broader RB-II literature), *only* the bidimensional model can yield criterion-level (85%) accuracy on either II subtask. It is therefore unsurprising that the responses of most II-II group participants performing at criterion were best-described by it.

Importantly, we do not wish to claim that DBM cannot yield relevant insights into participants' learning mechanisms (see Hélie, Turner, Crossley, Ell, & Ashby, 2017), nor do we intend to claim that self-reported strategies (which may certainly be subject to bias, as previously discussed) are a superior diagnostic tool. Instead, and for parsimony, we recommend interpreting humans' model fits with the same considerations and level of caution that we (and others) would typically extend to pigeons. That is, we caution against the assumption that a participant best fit by a specific decision bound is *necessarily* deploying a specific learning mechanism (for relevant considerations, see Edmunds, Wills, & Milton, 2025; Roark, Minda, Kalra, & Cruz, 2025).

Taken together, our data add to a growing body of evidence that RB and II tasks might not reliably engage disparate learning mechanisms among human learners (e.g., Carpenter, Wills, Benattayallah, & Milton, 2013; Edmunds et al., 2016; Newell et al., 2010). This conclusion would have clear implications for the ongoing debate surrounding the validity of RB-II dissociations (for reviews from each perspective, see Ashby & Valentin, 2017; Wills et al., 2019), as well as for the broader theoretical debate surrounding dual- vs. single-systems models of categorization (see Minda et al., 2024). Although our data cannot concretely resolve either issue, they do highlight the challenges inherent to the effort to isolate the contributions of associative mechanisms among human learners (McLaren et al., 2019), particularly in II tasks.

To this end, we also join Wills et al. (2019) in arguing that future research surrounding dual- vs. single-systems categorization would benefit from considering alternatives to standard RB and II paradigms that might allow for more reliable inferences surrounding participants' learning mechanisms (in particular, Wills and colleagues highlight the merits of the Shanks-Darby procedure; Shanks & Darby, 1998).

Do Declarative and Associative Mechanisms Produce Disparate Patterns of Task Switching Performance?

Of course, it is also possible that most of our human participants *did* arrive at the optimal strategies for each task condition, but that humans' declarative and associative mechanisms did not reliably produce different patterns of task switching performance in the context of the tasks considered here (Li et al., 2017). This possibility is at odds with past research suggesting that, among human learners, associative learning mechanisms may not produce switch costs (e.g., Forrest et al., 2012; Li et al., 2019a; Meier et al., 2013). However, this past research often involved tasks with small numbers of repeated stimuli, which are amenable to exemplar memorization.

Thus, the patterns of performance observed in prior studies of associative switching – characterized by small or no switch costs relative to large congruency effects (Forrest et al., 2012, Meier et al., 2013; Li et al., 2019a) – could reflect effortful memorization, which is arguably a declarative process (Edmunds et al., 2016; Smith et al., 2012; c.f. Bower & Winzenz, 1970; Forrest et al., 2012). By contrast, in the present project, we discouraged CSR memorization by instructing participants to learn about randomly-sampled stimuli through trial-and-error. Of course, if we then assume that our own II-II group participants were "true" associative learners (while noting the strong caveats discussed in the previous section), then our results beg the question as to why human associative learners might at least sometimes display switch costs when nonhuman associative learners rarely do so (Castro et al., 2016; Meier et al., 2013; Meier et al., 2016; O'Donoghue & Wasserman, 2021; see also Forrest et al., 2012, for proof-of-concept that associative mechanisms can produce small switch costs).

Of note, switch costs need not stem entirely from executive task set reconfiguration. As stressed by Logan and colleagues (Arrington & Logan, 2004; Logan & Bundesen, 2003; Schneider & Logan, 2005), many task switching paradigms (including ours) confound switching *tasks* with switching *task cues*. Logan and colleagues have argued that on stay trials, human participants may benefit from repetition priming of the task cue. This benefit is absent on switch trials, which thus incur switch costs. Nonetheless, subsequent research has demonstrated that executive task sets *do* play a substantial role in generating switch costs independently of repetition priming (Jost et al., 2013), and that even cue-switch costs might reflect the involvement of executive control processes (Grange & Houton, 2010).

In our paradigm, we cannot disentangle the effects of switching tasks from the effects of switching cues. It is possible that that the switch costs we observed among human RB-RB and II-II group participants, although similar in magnitude, may have differed in origin. For example, among the RB-RB group participants, switch costs may predominately reflect the influence of executive task sets, whereas among the II-II group participants, switch costs may predominately reflect repetition priming.

However, Meier et al. (2013) found that, when two cues were used to signal each of two RB subtasks, human associative learners (i.e., participants who were unable to report either subtask rule at the end of the experiment) did not display switch costs. As such, we have no compelling reason to suspect that cue switch costs would have been strongly present among our II-II group participants (although this remains an open question for future research).

Alternatively, it is also possible that the switch costs observed among our human II-II group participants might reflect the overarching structure of the task switching paradigm (such that task switching itself functions as a declarative task; we thank an anonymous reviewer for raising this possibility). In other words, even if II-II group participants were using the optimal associative strategies to solve each subtask, they might have nonetheless activated declarative task sets that incorporated the relevant CSR associations for a given subtask (for relevant discussion, see also Meier et al., 2013; Li et al., 2019). Of course, why such declarative task sets might have influenced performance in our experiments, but not in past research surrounding associative task switching using similar trial structures (e.g., Meier et al., 2013) remains to be determined.

Finally, it is also worth noting that the average age of our sample (mean = 43.10; *SD* = 11.95) skews slightly older than in past studies of associative task-switching. Although switch cost magnitudes do generally increase with age (Wasylyshyn, Verhaeghen, & Sliwinski, 2011), and age likewise slows overall learning speeds across RB and II tasks (alongside reducing the proportion of participants who converge on the optimal strategies according to DBM; e.g., Maddox, Pacheco, Reeves, Zhu, & Schnyer, 2010), it remains unclear whether these age-related changes might also influence the interaction between task condition (declarative vs. associative) and switch cost magnitudes.

Conclusions

Our data add to a growing body of research demonstrating that associative mechanisms can support robust task switching (Castro & Wasserman, 2016; Li et al., 2019a; Meier et al., 2013; Meier et al., 2016; O'Donoghue & Wasserman, 2022). In Experiment 1, we highlighted the challenges of isolating the contributions of associative mechanisms among human learners (particularly using II tasks), a critical issue that has rarely been acknowledged in studies of associative task switching (but see McLaren et al., 2019). Our data also raised the question as to whether associative learning mechanisms might still produce significant switch costs under at least some conditions (Forrest et al., 2012; Li et al., 2017).

Conversely, in Experiment 2, we demonstrated that pigeons – which presumably rely exclusively on associative mechanisms to solve both RB and II tasks (O'Donoghue et al., 2020; Smith et al., 2012) – display *no* switch costs in either task condition. These data highlight the impressive power and flexibility that associative mechanisms can afford (see also Lind, 2019; McMurray et al., 2012; Sloutsky et al., 2017; Wasserman et al., 2023), despite the fact that they are often overlooked in discussions of "complex" cognitive processes.

Overall, our data point to three clear directions for future research. First, more work is needed to disentangle the contributions of declarative and associative mechanisms in RB- and II-learning (see also Edmunds et al., 2016; Edmunds et al., 2018; Wills et al., 2019). Second, future research involving alternative paradigms could help determine whether human learners reliably do produce switch costs in associative tasks wherein CSR memorization in precluded. Finally, we look forward to continued research exploring the striking flexibility afforded by associative mechanisms (Turner & Wasserman, 2023). Given the difficulties of isolating associative mechanisms among human learners (McLaren et al., 2019), we suspect that comparative investigations will prove especially fruitful in this domain.

Acknowledgements

The experiments reported here were previously reported in the lead author’s dissertation (O’Donoghue, 2022). We thank Rey Bernhard and Hanlong Fan for their work coding participants’ self-reports in Experiment 1.

Open Practices Statement

The data and materials for Experiments 1 and 2 are available via https://osf.io/3265r/ (O'Donoghue, 2023). Neither experiment was preregistered.

References

Allen, C., & Bekoff, M. (1995). Cognitive ethology and the intentionality of animal behaviour. *Mind & Language, 10*(4), 313-328. https://doi.org/10.1111/j.1468-0017.1995.tb00017.x

Ardiale, E., Hodzik, S., & Lemaire, P. (2012). Aging and strategy switch costs: A study in arithmetic problem solving. *L'Année Psychologique*, *112*(3), 345-360. https://doi.org/10.3917/anpsy.123.0345

Arrington, C. M., & Logan, G. D. (2004). Episodic and semantic components of the compound stimulus strategy in the explicit task-cuing procedure. *Memory & Cognition, 32*(6), 965- 978. https://doi.org/10.3758/BF03196874

Ashby, F. G., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442-481. https://doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 33–53. https://doi.org/10.1037/0278-7393.14.1.33

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95*(1), 124–150. https://doi.org/10.1037/0033-295X.95.1.124

Ashby, F. G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory*. Psychological Review, 124*(4), 472–482. https://doi.org/10.1037/rev0000064

Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. In *Handbook of Categorization in Cognitive Science (Second Edition).* H. Cohen & C. Lefebvre, Eds.: 157-188. Elsevier. New York. https://doi.org/10.1016/B978-0-08-101107-2.00007-5

Ashby, F. G., & Valentin, V. V. (2018). The categorization experiment: Experimental design and data analysis. In *Stevens Handbook of Experimental Psychology and Cognitive Neuroscience, Fourth Edition, Volume Five: Methodology.* Wagenmakers, E. & Wixted, J.T., Eds.: 307-333. Wiley. New York.

Bower, G. H., & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science*, *20*(2), 119-120. https://doi.org/10.3758/BF03335632

Broschard, M. B., Kim, J., Love, B. C., Wasserman, E. A., & Freeman, J. H. (2019). Selective attention in rat visual category learning. *Learning & Memory*, *26*(3), 84-92. https://doi.org/ 10.1101/lm.048942.118

Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition*, *40*(3), 434-449. https://doi.org/10.3758/s13421-011-0154-4

Castro, L., & Wasserman, E. A. (2016). Executive control and task switching in pigeons. *Cognition, 146*, 121-135. https://doi.org/10.1016/j.cognition.2015.07.014

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. *Frontiers in Psychology*, *5*, 825. https://doi.org/10.3389/fpsyg.2014.00825

Crossley, M. J., Roeder, J. L., Helie, S., & Ashby, F. G. (2018). Trial-by-trial switching between procedural and declarative categorization systems. *Psychological Research*, *82*(2), 371-384. https://doi.org/10.1007/s00426-016-0828-4

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 933-948. https://doi.org/10.1037/xlm0000083

Dreisbach, G., & Haider, H. (2008). That's what task sets are for: shielding against irrelevant information. *Psychological Research*, *72*(4), 355-361. https://doi.org/10.1007/s00426-007-0131-5

Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly Journal of Experimental Psychology*, *68*, 1203-1222. https://doi.org/10.1080/17470218.2014.978875

Edmunds, C. E., Milton, F., & Wills, A. J. (2018). Due process in dual process: Model-recovery simulations of decision-bound strategy analysis in category learning. *Cognitive Science*, *42*, 833-860. https://doi.org/10.1111/cogs.12607

Edmunds, C. E. R., Wills, A. J., & Milton, F. N. (2016). Memory for exemplars in category learning. In A. Papafragou, D. Grodner, D. Mirman & J. C. Truswell (Eds.), Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 2243–2248). Austin, TX: Cognitive Science Society.

Edmunds, C. E. R., Wills, A. J., & Milton, F. (2019). Initial training with difficult items does not facilitate category learning. *Quarterly Journal of Experimental Psychology*, *72*(2), 151-167. https://doi.org/10.1080/17470218.2017.1370477

Edmunds, C. E. R., Wills, A. J., & Milton, F. (2025). Structure-based dissociations provide agnostic evidence to the multiple-systems debate. *Nature Reviews Psychology, 4*(1), 66-66. https://doi.org/10.1038/s44159-024-00395-w

Ellefson, M. R., Shapiro, L. R., & Chater, N. (2006). Asymmetrical switch costs in children. *Cognitive Development*, *21*(2), 108-130. https://doi.org/10.1016/j.cogdev.2006.01.002

Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, *36*(4), 749-761. https://doi.org/10.3758/MC.36.4.749

Forrest, C. L., Elchlepp, H., Monsell, S., & McLaren, I. P. L. (2012, August). Task switching without knowledge of the tasks. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 1548–1553). Austin, TX: Cognitive Science Society.

García-Patos, P., & Olmos, R. (2020). Multiple imputation in multilevel models. A revision of the current software and usage examples for researchers. *The Spanish Journal of Psychology, 23*, 1-19. https://doi.org/10.1017/SJP.2020.48

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*(3), 297-337. https://doi.org/10.1006/cogp.2001.0770

Gopher, D., Armony, L., & Greenshpan, Y. (2000). Switching tasks and attention policies. *Journal of Experimental Psychology: General, 129*(3), 308–339. https://doi.org/10.1037/0096-3445.129.3.308

Grange, J. A., & Houghton, G. (2010). Cue-switch costs in task-switching: cue priming or control processes? *Psychological Research, 74*(5), 481-490. https://doi.org/10.1007/s00426-009-0270-y

Hélie, S., Turner, B. O., Crossley, M. J., Ell, S. W., & Ashby, F. G. (2017). Trial-by-trial identification of categorization strategy using iterative decision-bound modeling. *Behavior Research Methods*, *49*(3), 1146-1162. https://doi.org/10.3758/s13428-016-0774-5

Herbranson, W. T., Karas, E., & Hardin, G. (2017). Perception of angle in visual categorization by pigeons (Columba livia). *Animal Behavior and Cognition, 4(3)*, 286-300. https://doi.org/10.26451/abc.04.03.07.2017

Holland, P. C. (1985). The nature of conditioned inhibition in serial and simultaneous feature negative discriminations. In R. R. Miller & N. E. Spear (Eds.), *Information Processing in Animals: Conditioned Inhibition* (pp. 267-297). Hillsdale, NJ: Erlbaum.

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, *46*(3), 702-721. https://doi.org/10.3758/s13428-013-0411-5

Jost, K., De Baene, W., Koch, I., & Brass, M. (2015). A review of the role of cue processing in task switching. *Zeitschrift für Psychologie, 221*(1), 5-14. https://doi.org/10.1027/2151-2604/a000125

Kelleher, R. T., Riddle, W. C., & Cook, L. (1962). Observing responses in pigeons. *Journal of the Experimental Analysis of Behavior*, *5*(1), 3-13. https://doi.org/10.1901/jeab.1962.5-3

Kiesel, A., Wendt, M., & Peters, A. (2007). Task switching: On the origin of response congruency effects. *Psychological Research*, *71*(2), 117-125. https://doi.org/10.1007/s00426-005-0004-8

Kleiner, M., Brainard, D., & Pelli, D., 2007. What’s new in Psychtoolbox-3? *Perception, 36*, ECVP Abstract Supplement.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, *135*(2), 162-183. https://doi.org/10.1037/0096-3445.135.2.162

Lea, S. E. G., Wills, A. J., Leaver, L. A., Ryan, C. M. E., Bryant, C. M. L., & Millar, L. (2009). A comparative analysis of the categorization of multidimensional stimuli: II. Strategic information search in humans (*Homo sapiens*) but not in pigeons (*Columba livia*). *Journal of Comparative Psychology, 123*(4), 406–420. https://doi.org/10.1037/a0016851

Li, X., Li, B., Lages, M., & Stoet, G. (2017). Commentary: task-switching in pigeons: associative learning or executive control?. *Frontiers in Psychology*, *8*, 1420. https://doi.org/10.3389/fpsyg.2017.01420

Li, B., Li, X., Liu, X., Lages, M., & Stoet, G. (2019a). Target-response associations can produce response-congruency effects without task-switching costs. *Frontiers in Psychology*, *10*, 40. https://doi.org/10.3389/fpsyg.2019.00040

Li, X., Li, B., Liu, X., Lages, M., & Stoet, G. (2019b). Task-switching costs disappear if non-Chinese participants respond to Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(11), 2051–2071. https://doi.org/10.1037/xlm0000692

Lind, J. (2018). What can associative learning do for planning?. *Royal Society Open Science*, *5*(11), 180778. https://doi.org/10.1098/rsos.180778

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29*(3), 575–599. https://doi.org/10.1037/0096-1523.29.3.575

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural processes*, *66*(3), 309-332. https://doi.org/10.1016/j.beproc.2004.03.011

Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Rule-based and information-integration category learning in normal aging. *Neuropsychologia*, *48*(10), 2998-3008. https://doi.org/10.1016/j.neuropsychologia.2010.06.008

Maes, E., De Filippo, G., Inkster, A. B., Lea, S. E., De Houwer, J., D'Hooge, R., ... & Wills, A. J. (2015). Feature-versus rule-based generalization in rats, pigeons and humans. *Animal Cognition*, *18*(6), 1267-1284. https://doi.org/10.1007/s10071-015-0895-8

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77-80. https://doi.org/10.1126/science.283.5398.77

Matsuki, K. (2017). grt: General Recognition Theory. R package version 0.2.1. https://CRAN.R-project.org/package=grt

McLaren, I. P. L., McAndrew, A., Angerer, K., McLaren, R., Forrest, C., Bowditch, W., ... & Verbruggen, F. (2019). Mackintosh Lecture—: Association and cognition: Two processes, one system. *Quarterly Journal of Experimental Psychology, 72*(2), 98-117. https://doi.org/10.1177/1747021818766287

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831-877. https://doi.org/10.1037/a0029872.

Meier C, Lea SE, Forrest CL, Angerer K, & McLaren IP (2013). Comparative evidence for associative learning in task switching. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1020 –1025). Austin, TX: Cognitive Science Society.

Meier, C., Lea, S. E. G., & McLaren, I. P. L. (2016). Task-switching in pigeons: Associative learning or executive control? *Journal of Experimental Psychology: Animal Learning and Cognition, 42*(2), 163–176. https://doi.org/10.1037/xan0000100

Minda, J. P., Roark, C. L., Kalra, P., & Cruz, A. (2024). Single and multiple systems in categorization and category learning. *Nature Reviews Psychology*, *3*(8), 536-551. https://doi.org/10.1038/s44159-024-00336-7

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

Mulcahy, N. J., & Call, J. (2006). Apes save tools for future use. *Science, 312*(5776), 1038-1040. https://doi.org/10.1126/science.1125456

Navarro, V. M., Jani, R., & Wasserman, E. A. (2019). Pigeon category learning: Revisiting the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: Animal Learning and Cognition, 45*(2), 174–184. https://doi.org/10.1037/xan0000198

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 199-203. https://doi.org/10.1002/wics.199

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*(5), 563-581. https://doi.org/10.3758/MC.38.5.563

Newell, B. R., Moore, C. P., Wills, A. J., & Milton, F. (2013). Reinstating the frontal lobes? Having more time to think improves implicit perceptual categorization: A comment on Filoteo, Lauritzen, and Maddox (2010). *Psychological Science*, *24*(3), 386-389. https://doi.org/10.1177/0956797612457387

Nieuwenhuis, S., & Monsell, S. (2002). Residual costs in task switching: Testing the failure-to-engage hypothesis. *Psychonomic Bulletin & Review, 9*(1), 86-92. https://doi.org/10.3758/BF03196259

Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity?. *Memory & Cognition*, *33*(7), 1256-1271. https://doi.org/10.3758/BF03193227

O'Donoghue, E. M. (2022). A Comparative Investigation into the Mechanisms that Support Multidimensional Categorization (Doctoral dissertation, The University of Iowa).

O’Donoghue, E.M. (2023). RB-II Task-Switching. Retrieved from osf.io/3265r

O'Donoghue, E. M., Broschard, M. B., Freeman, J. H., & Wasserman, E. A. (2022). The Lords of the Rings: People and pigeons take different paths mastering the concentric-rings categorization task. *Cognition*, *218*, 104920. https://doi.org/10.1016/j.cognition.2021.104920

O'Donoghue, E. M., Broschard, M. B., & Wasserman, E. A. (2020). Pigeons exhibit flexibility but not rule formation in dimensional learning, stimulus generalization, and task switching. *Journal of Experimental Psychology: Animal Learning and Cognition*, *46*(2), 107-123. https://doi.org/10.1037/xan0000234

O'Donoghue, E. M., Castro, L., & Wasserman, E. A. (2022). Hierarchical and configural control in conditional discrimination learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 48*(4), 370–382. https://doi.org/10.1037/xan0000342

O'Donoghue, E., & Wasserman, E. A. (2021). Pigeons proficiently switch among four tasks without cost. *Journal of Experimental Psychology: Animal Learning and Cognition*, *47*(2), 150–162. https://doi.org/10.1037/xan0000287

Pontes, A. C., Mobley, R. B., Ofria, C., Adami, C., & Dyer, F. C. (2020). The evolutionary origin of associative learning. *The American Naturalist, 195*(1), E1-E19. https://doi.org/10.5061/dryad.f45gh6s

Roark, C. L., Minda, J. P., Kalra, P., & Cruz, A. (2025). Reply to 'Structure-based dissociations provide agnostic evidence to the multiple-systems debate.' *Nature Reviews Psychology, 4*(1), 67-67. https://doi.org/10.1038/s44159-024-00396-9

Samavatyan, H., & Leth-Steensen, C. (2009). The time course of task switching: A speed—accuracy trade-off analysis. *Memory & Cognition*, *37*(7), 1051-1058. https://doi.org/10.3758/MC.37.7.1051

Schneider, D. W., & Logan, G. D. (2005). Modeling Task Switching Without Switching Tasks: A Short-Term Priming Account of Explicitly Cued Performance. *Journal of Experimental Psychology: General, 134*(3), 343–367. https://doi.org/10.1037/0096-3445.134.3.343

Shanks, D. R., & Darby, R. J. (1998). Feature-and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 24*(4), 405-415. https://doi.org/10.1037/0097-7403.24.4.405

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484-489. https://doi.org/10.1038/nature16961

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Hassabis, D. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv.* https://doi.org/10.48550/arXiv.1712.01815

Slama, H., Rebillon, E., & Kolinsky, R. (2017). Expertise and cognitive flexibility: A musician's tale. *Journal of Cultural Cognitive Science*, *1*(2), 119-127. https://doi.org/10.1007/s41809-017-0011-5

Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology*, *97*, 1-30. https://doi.org/10.1016/j.cogpsych.2017.06.001

Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review, 18*, 414-421. https://doi.org/10.3758/s13423-010-0047-8

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., ... & Grace, R. C. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2355-2369. https://doi.org/10.1016/j.neubiorev.2012.09.003

Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science*, *25*(2), 447-457. https://doi.org/10.1177/0956797613509112

Stephens, R. G., & Kalish, M. L. (2018). The effect of feedback delay on perceptual category learning and item memory: Further limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1397-1413. https://doi.org/10.1037/xlm0000528

Stoet, G., & Snyder, L. H. (2007). Extensive practice does not eliminate human switch costs. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 192-197. https://doi.org/10.3758/CABN.7.3.192

Turner, B. M., & Wasserman, E. A. (2023). The pigeon as a machine: Complex category structures can be acquired by a simple associative model. *iScience, 26*(10), 107998. https://doi.org/10.1016/j.isci.2023.107998

Wasserman, E. A., Kain, A. G., & O'Donoghue, E. M. (2023). Resolving the associative learning paradox by category learning in pigeons. *Current Biology, 33*(6), 1112-1116. https://doi.org/10.1016/j.cub.2023.01.024

Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: a meta-analysis. *Psychology and Aging*, *26*(1), 15-20. https://doi.org/10.1037/a0020912

Weinberger, A. B., & Green, A. E. (2022). Dynamic development of intuitions and explicit knowledge during implicit learning. *Cognition, 222*, 105008. https://doi.org/10.1016/j.cognition.2021.105008

Wills, A.J., Edmunds, C.E.R., Le Pelley, M.E., Milton, F., Newell, B.R., Dwyer, D.M., & Shanks, D.R. (2019). Dissociable learning processes, associative theory, and testimonial reviews: A comment on Smith and Church (2018). *Psychonomic Bulletin & Review, 26*, 1988–1993. https://doi.org/10.3758/s13423-019-01644-3

Wills, A. J., Lea, S. E., Leaver, L. A., Osthaus, B., Ryan, C. M., Suret, M. B., ... & Millar, L. (2009). A comparative analysis of the categorization of multidimensional stimuli: I. Unidimensional classification does not necessarily imply analytic processing; evidence from pigeons (Columba livia), squirrels (Sciurus carolinensis), and humans (Homo sapiens). *Journal of Comparative Psychology*, *123*(4), 391-405. https://doi.org/10.1037/a001621

Zaki, S. R., & Kleinschmidt, D. F. (2014). Procedural memory effects in categorization: Evidence for multiple systems or task complexity?. *Memory & Cognition*, *42*(3), 508-524. https://doi.org/10.3758/s13421-013-0375-9

Zentall, T. R. (2020). The midsession reversal task: A theoretical analysis. *Learning & Behavior*, *48*, 195-207. https://doi.org/10.3758/s13420-020-00423-8